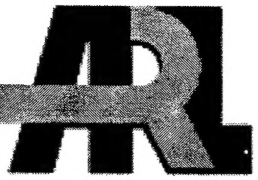
ARMY RESEARCH LABORATORY



Frequency-Hopping Communication Systems

by Don Torrieri

ARL-TR-2949 March 2003

NOTICES

Disclaimers

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.

Army Research Laboratory
Adelphi, MD 20783-1197

ARL-TR-2949 **March 2003**

Frequency-Hopping Communication Systems

Don Torrieri
Computational and Information Sciences Directorate, ARL

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED (From - To)
March 2003	Final	

4. TITLE AND SUBTITLE
Frequency-Hopping Communication Systems

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Don Torrieri

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Research Laboratory
ATTN: AMSRL-CI-C
2800 Powder Mill Road
Adelphi, MD 20783-1197

8. PERFORMING ORGANIZATION REPORT NUMBER
ARL-TR-2949

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Laboratory
2800 Powder Mill Road
Adelphi, MD 20783-1197

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This report presents a current review and analysis of the most important aspects of frequency-hopping communication systems. Frequency synthesis and other implementation issues are discussed. The effects and suitability of various modulations is examined. The impact of error-correcting codes on the performance of frequency-hopping systems is analyzed.

15. SUBJECT TERMS
Frequency hopping, FH/CPFSK, FH/MSK

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Don Torrieri
Unclassified	Unclassified	Unclassified	UL	186	19b. TELEPHONE NUMBER *(Include area code)* 301-394-2484

Preface

This report explains and derives the fundamentals of frequency-hopping communication systems. The level of presentation is suitable for those with a solid background in the theory of digital communications. Throughout the report, there are many streamlined derivations, new derivations, and simplications of the classical theory. Section 1 provides the basic definitions, concepts, and rationale of frequency hopping. Section 2 presents an overview and comparison of direct, digital, and indirect frequency synthesizers. Section 3, 4, and 5 concentrate on a comparison of the modulations that might be used in a frequency-hopping system. Both spectral characteristics and the influence on system performance are examined. The effectiveness of Reed-Solomon codes, trellis-coded modulation, and turbo codes in suppressing partial-band interference is analyzed in Section 6. Section 7 treats frequency-hopping code-division multiple-access (CDMA) for both peer-to-peer and cellular systems. The advantages of frequency hopping in network applications and the effects of spatial diversity, spectral splatter, and the number of equivalent frequency channels are explained. Section 8 considers issues in the synchronization of frequency-hopping patterns. The comparison and combination of frequency-hopping and direct-sequence systems are discussed in Section 9. Section 10 examines the detection of frequency-hopping signals with an emphasis on the use of radiometers. Because of the essential role of error- correcting codes in frequency-hopping communication systems, some of the fundamental results of coding theory are presented in Appendix A and used to derive the appropriate receiver computations and the error probabilities at receiver outputs. Appendix B provides derivations of the signal representations used throughout the main text.

Contents

Figures

Tables

1. Definitions and Concepts

A *spread-spectrum signal* is a signal that has an extra modulation in addition to the data modulation so that its bandwidth is much wider than is required by the underlying data-modulated signal. Spread-spectrum communication systems are useful for suppressing interference, making interception difficult, accommodating fading and multipath channels, and providing a multiple-access capability. The most practical and dominant methods of spread-spectrum communications are direct-sequence modulation and frequency hopping of digital communications.

Frequency hopping is the periodic changing of the carrier frequency of a transmitted signal. The sequence of carrier frequencies is called the *frequency-hopping pattern*. The set of M possible carrier frequencies $\{f_1, f_2, \ldots, f_M\}$ is called the *hopset*. Hopping occurs over a frequency band called the *hopping band* that includes M *frequency channels*. Each frequency channel is defined as a spectral region that includes a single carrier frequency of the hopset as its center frequency and has a bandwidth B large enough to include most of the power in a signal pulse with the carrier frequency. Figure 1 illustrates the frequency channels associated with a particular frequency-hopping pattern. The time interval between hops is called the *hop interval*, and its duration is called the *hop duration* or the *hopping period* and is denoted by T_h. The hopping band has bandwidth $W \geq MB$.

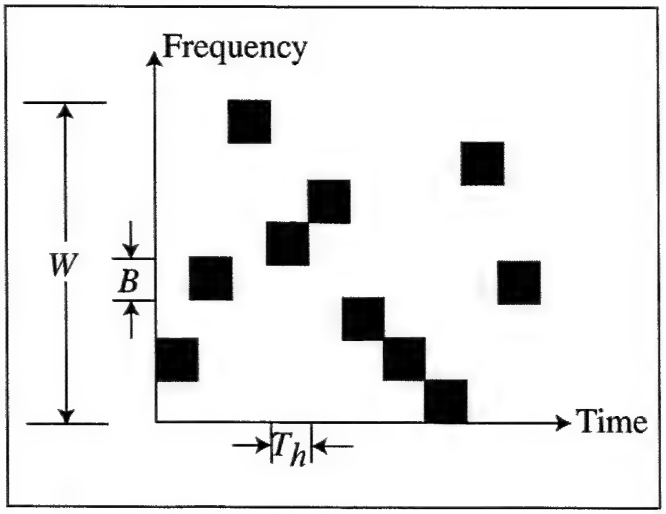


Figure 1. Frequency-hopping patterns.

Figure 2 depicts the general form of a frequency-hopping system. The frequency synthesizers produce frequency-hopping patterns determined by the time-varying multilevel sequence specified by the output bits of the code generators. In the transmitter, the data-modulated signal is mixed with the synthesizer output pattern to produce the frequency-hopping signal. If the data modulation is some form of angle modulation $\phi(t)$, then the received signal for the ith hop is

$$s(t) = \sqrt{2S}\cos\left[2\pi f_j t + \phi(t) + \phi_i\right], \quad (i-1)T_h \le t \le iT_h \qquad (1\text{-}1)$$

where S is the average power, f_j is the carrier frequency for this hop, and ϕ_i is a random phase angle for the ith hop. The frequency-hopping pattern produced by the receiver synthesizer is synchronized with the pattern produced by the transmitter, but is offset by a fixed intermediate frequency, which may be zero. The mixing operation removes the frequency-hopping pattern from the received signal. The mixer output is applied to a bandpass filter that excludes double-frequency components and power that originated outside the appropriate frequency channel and produces the data-modulated *dehopped signal*, which has the form of (1-1) with f_j replaced by the intermediate frequency.

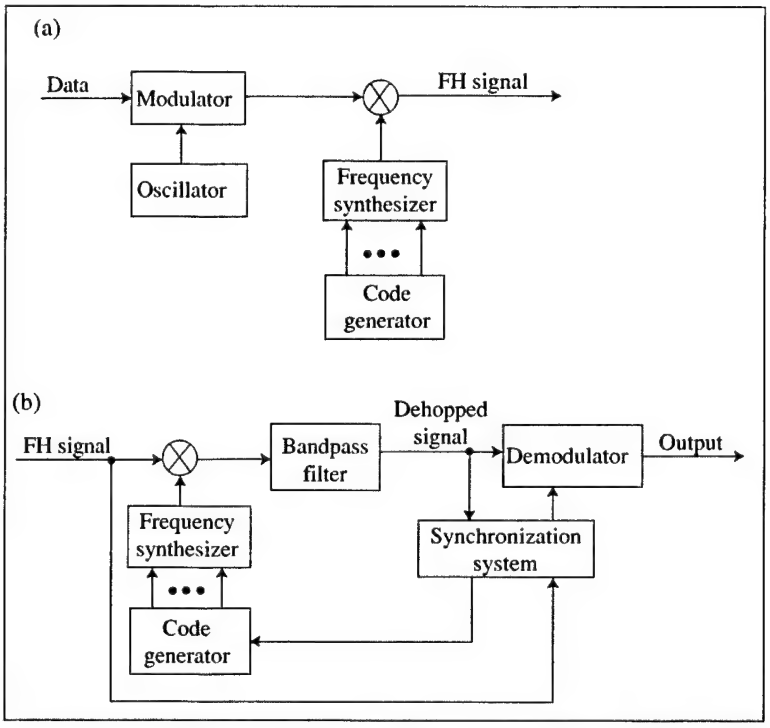


Figure 2. General form of frequency-hopping system: (a) transmitter and (b) receiver.

Although it provides no advantage against white noise, frequency hopping enables signals to hop out of frequency channels with interference or slow frequency-selective fading. To fully exploit this capability against narrowband interference signals, disjoint frequency channels are necessary. The disjoint channels may be contiguous or have unused spectral regions between them.

Some spectral regions with steady interference or a susceptibility to fading may be omitted from the hopset, a process called *spectral notching*. Multiple frequency-shift keying (MFSK) differs fundamentally from frequency hopping in that all the MFSK subchannels affect each receiver decision. No escape from or avoidance of a subchannel with interference is possible.

To ensure the secrecy and unpredictability of the frequency-hopping pattern, the pattern should be a pseudorandom sequence of frequencies with a large period and a uniform distribution over the frequency channels, and should be generated by a multilevel sequence with a large linear span. The large period prevents the capture and storage of a period of the pattern by an opponent. The *linear span* of a multilevel sequence is the smallest degree of any linear recursion that the sequence satisfies. A large linear span inhibits the reconstruction of the pattern from a short segment of it. The set of output bits produced by the code generator usually constitutes a symbol drawn from a finite field with the necessary properties. A frequency-hopping pattern is obtained by associating a distinct frequency with each symbol. A number of methods have been found to ensure a large linear span [1], [2].

An architecture that enhances the security of synthesizer control is shown in Figure 3. The specific algorithm for generating the control bits is determined by the key and the time-of-day (TOD). The *key* is a set of bits that are changed infrequently and must be kept secret. The TOD is a set of bits that are derived from the stages of the TOD counter and change with every transition of the TOD clock. For example, the key might change daily while the TOD might change every second. The purpose of the TOD is to vary the generator algorithm without

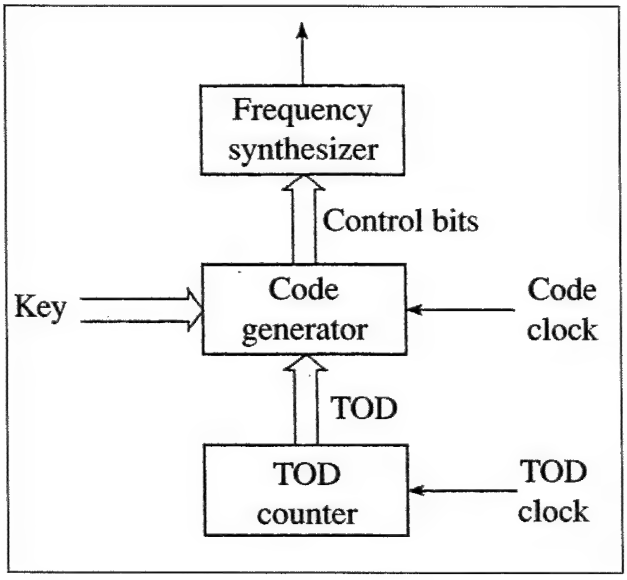


Figure 3. Secure method of synthesizer control.

constantly changing the key. In effect, the generator algorithm is controlled by a time-varying key. The code clock, which regulates the changes of state in the code generator and thereby controls the hopping rate, operates at a much higher rate than the TOD clock. In a receiver, the code clock is produced by the synchronization system. In both the transmitter and the receiver, the TOD clock may be derived from the code clock.

A frequency-hopping pulse with a fixed carrier frequency occurs during a portion of the hop interval called the *dwell interval*. As illustrated in Figure 4, the *dwell time* is the duration of the dwell interval during which the channel symbols are transmitted. The hop duration T_h is equal to the sum of the dwell time T_d and the switching time T_{sw}. The *switching time* is equal to the *dead time*, which is the duration of the interval when no signal is present, plus the rise and fall times of a pulse. Even if the switching time is absent in the transmitted signal, it will be present in the dehopped signal in the receiver because of the imperfect synchronization of received and receiver-generated waveforms. The nonzero switching time, which may include an intentional *guard time*, decreases the transmitted symbol duration T_s. If T_{so} is the symbol duration in the absence of frequency hopping, then $T_s = T_{so}(T_d/T_h)$. The reduction in symbol duration expands the transmitted spectrum and thereby reduces the number of frequency channels within a fixed hopping band. Since the receiver filtering will ensure that rise and fall times of pulses have durations on the order of a symbol duration, $T_{sw} > T_s$ in practical systems.

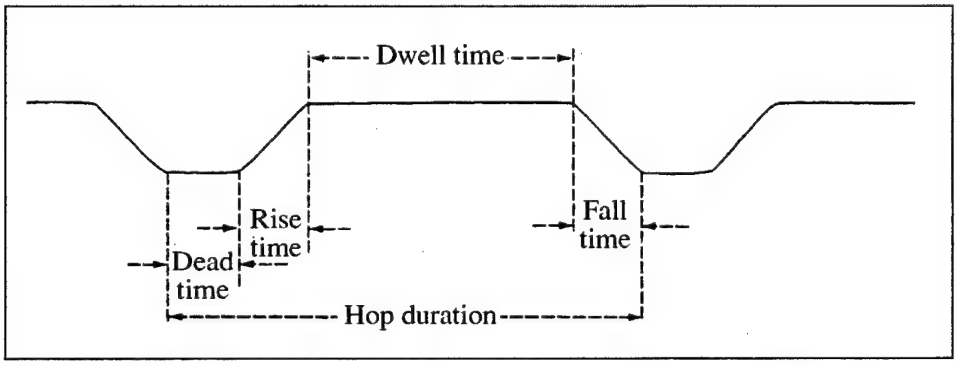


Figure 4. Time durations of a frequency-hopping pulse.

2. Frequency Synthesizers

A *frequency synthesizer* [3], [4], [5], [6] in a frequency-hopping system converts a stable reference frequency into the various frequencies of the hopset. The *reference signal*, which is a tone at the reference frequency, is usually the output of a frequency divider or multiplier fed by a stable frequency source, such as an atomic or crystal oscillator. The use of a single reference signal ensures that any output frequency of the synthesizer has the same stability and accuracy as the reference. The three fundamental types of frequency synthesizers are the direct, digital, and indirect synthesizers. Most practical synthesizers are hybrids of these fundamental types.

2.1 Direct Frequency Synthesizer

A *direct frequency synthesizer* uses frequency multipliers and dividers, mixers, bandpass filters, and electronic switches to produce output signals at the desired frequencies. Among the many methods of direct synthesis, the standard approach is the *double-mix-divide* technique illustrated in Figure 5. The reference signal at frequency f_r is mixed with a tone at the fixed frequency f_a. The bandpass filter selects the sum frequency $f_r + f_a$ produced by the mixer. Another mixing and filtering operation with a tone at $f_b + f_1$ produces the frequency $f_r + f_a +$ $f_b + f_1$. If the fixed frequencies f_a and f_b are chosen so that

$$f_a + f_b = 9f_r \tag{2-1}$$

then the divider produces the output frequency $f_r + f_1/10$. In principle, a single mixer and bandpass filter could produce this output frequency, but two mixers and bandpass filters simplify the filters. Each bandpass filter must select the sum frequency while suppressing the difference frequency and the mixer input frequencies, which may enter the filter because of mixer leakage. If the sum frequency is too close to one of these other frequencies, the bandpass filter becomes prohibitively complex and expensive.

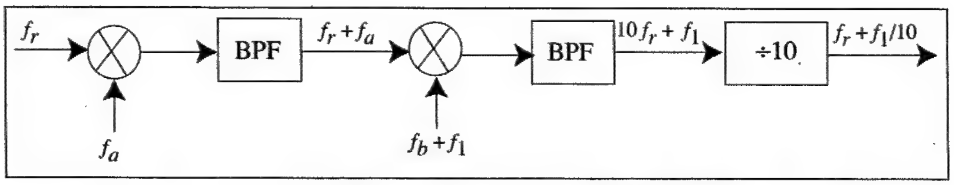


Figure 5. Double-mix-divide module, where BPF = bandpass filter.

The double-mix-divide (DMD) system of Figure 5 can be used as a module in a direct frequency synthesizer that can achieve arbitrary frequency resolution by cascading enough DMD modules. A synthesizer that provides two-digit resolution is shown in Figure 6. When the synthesizer is used in a frequency-hopping system, the control bits are produced by the code generator. Each decade switch passes a single tone to a DMD module. The ten tones that are available to the decade switches may be produced by applying the reference frequency to appropriate frequency multipliers and dividers in the tone generator. Equation (2-1) ensures that the output frequency of the second bandpass filter in DMD module 2 is $10f_r + f_2 + f_1/10$. Thus, the final synthesizer output frequency is $f_r + f_2/10 + f_1/100$.

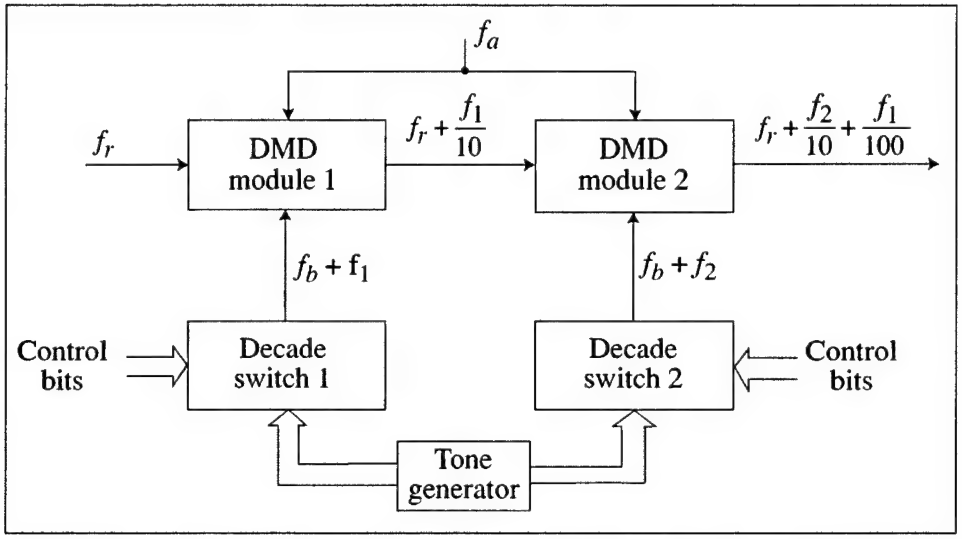


Figure 6. Direct frequency synthesizer with two-digit resolution.

Example 1. It is desired to produce a 1.79 MHz tone. Let f_r = 1 MHz and f_b = 5 MHz. The ten tones provided to the decade switches are 5, 6, 7, . . . , 14 MHz so that f_1 and f_2 can range from 0 to 9 MHz. Equation (2-1) yields f_a = 4 MHz. If f_1 = 7 MHz and f_2 = 9 MHz, then the output frequency is 1.79 MHz. The frequencies f_a and f_b are such that the designs of the bandpass filters inside the modules are reasonably simple.□

Direct frequency synthesizers provide both very fine resolution and high frequencies, but often require a very large amount of hardware. They do not provide a phase-continuous output after frequency switching.

2.2 Digital Frequency Synthesizer

A *digital frequency synthesizer* converts the stored sample values of a sine wave into an analog sine wave with a specified frequency. The periodic and symmetric character of a sine wave implies that only values for the first quadrant need to be stored. The basic elements of a digital frequency synthesizer are shown in

Figure 7. The frequency data, which is produced by the code generator in a frequency- hopping system, comprises a set of bits that determine the synthesized frequency by specifying a phase increment δ. The accumulator converts the phase increment into successive samples of the phase by adding the increment to the content of an internal register after every cycle of the reference signal. A phase sample $\theta = n\delta, n = 0, 1, \ldots,$ defines an address in the read-only memory (ROM) at which the value $\sin\theta$ is stored. This value is applied to a digital-to-analog converter (DAC), which performs a sample-and-hold operation. The converter output is filtered to produce the desired analog signal.

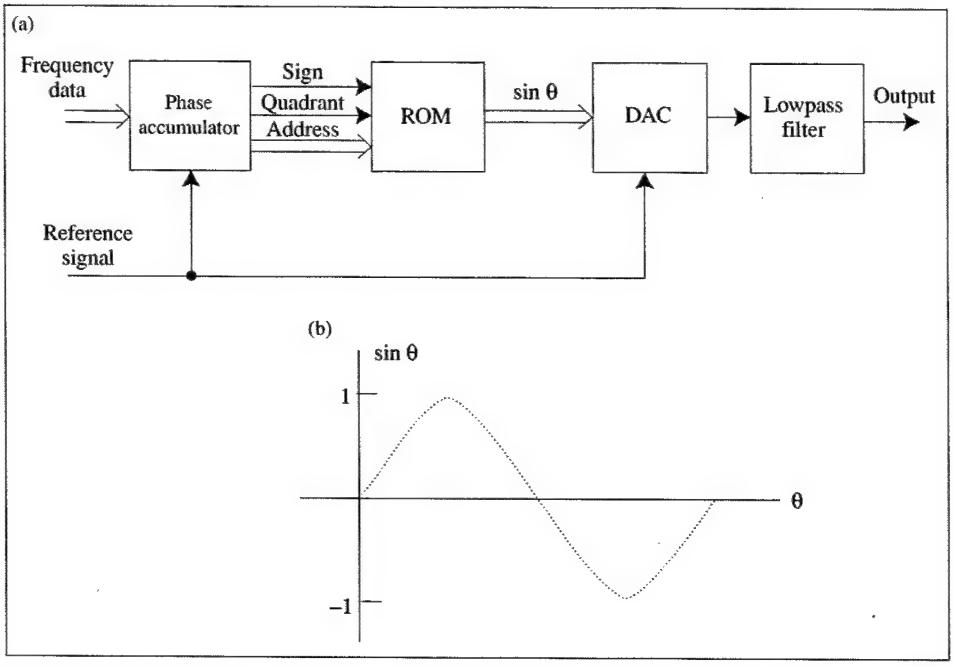


Figure 7. (a) Digital frequency synthesizer and (b) sample values of $\sin\theta$ over one period.

Let N denote the numerical capacity of the accumulator, which determines the maximum number of ROM addresses that the phase accumulator can specify. One sample value of $\sin\theta$ is read out after every cycle of the reference signal. If N sample values are read out during each period of $\sin\theta$, then the frequency of the analog signal produced is

$$\Delta = \frac{f_r}{N} \tag{2-2}$$

where f_r is the frequency of the reference signal. The output frequency f_{0k} is produced when every kth stored sample value is read out at the reference rate. Thus, if the phase accumulator increments by k after every cycle of the reference signal, then

$$f_{0k} = k\Delta \tag{2-3}$$

which implies that Δ is the frequency resolution and the minimum frequency that can be generated by the synthesizer.

The maximum frequency f_m that can be generated is produced by using only a few samples of sin θ per period. From the Nyquist sampling theorem, it is known that $f_m < f_r/2$ is required to avoid aliasing. Practical DAC and lowpass filter requirements further limit f_m to approximately 0.4 f_r or less. Thus, $q \geq 2.5$ samples of sin θ per period are used in synthesizing f_m, and

$$f_m = \frac{f_r}{q} \tag{2-4}$$

The lowpass filter may be implemented with a linear phase across a flat passband extending to approximately f_m. The frequencies f_r and f_m are limited by the speed of the digital-to-analog converter.

Let ν denote the number of bits in the accumulator register. The numerical capacity of the accumulator is $N = 2^\nu$. Suppose that $f_{\min}$ and $f_{\max}$ are specified minimum and maximum frequencies that must be produced by a synthesizer. Equations (2-2) and (2-3) imply that $f_{\min} \geq f_r/N$ and $f_{\max} \leq f_r/q$. Therefore, $qf_{\max}/f_{\min} \leq N = 2^\nu$ and the required number of accumulator bits is

$$\nu = \lfloor \log_2(qf_{\max}/f_{\min}) \rfloor + 1 \tag{2-5}$$

where $\lfloor x \rfloor$ denotes the largest integer in x.

Suppose that 2^n distinct n-bit words are stored in the ROM. Each word represents one possible value of sin θ in the first quadrant or, equivalently, one possible magnitude of sin θ. The input to the ROM comprises $n + 2$ parallel bits. The two most significant bits are the *sign bit* and the *quadrant bit*. The sign bit specifies the polarity of sin θ. The quadrant bit specifies whether sin θ is in the first or second quadrants or in the third or fourth quadrants. The n least significant bits of the input determine the address in which the value of sin θ is stored when θ is in the first or third quadrants. The address specified by the n least significant bits is appropriately modified by the quadrant bit when θ is in the second or fourth quadrants. The sign bit becomes one of the $n + 1$ ROM output bits. Since the phase accumulator uses $\nu \geq n + 2$ bits, and $n + 2$ bits are needed by the ROM, the $\nu - n - 2$ least significant bits of the accumulator output are not applied to the ROM. The memory requirements of a ROM and the number of its input bits can be reduced by using trigonometric identities and hardware multipliers.

Since n ROM output bits specify the magnitude of sin θ, the quantization error produces the worst-case noise power

$$E_q = (2^{-n})^2 = -6n \text{ dB} \tag{2-6}$$

in the digital-to-analog converter output. The magnitude of E_q is called the *spectral purity* of the synthesizer.

Example 2. It is desired to design a digital synthesizer to cover 1 Hz to 1 MHz with a spectral purity greater than 45 dB. According to (2-6), the use of 8-bit words in the ROM is adequate for the required spectral purity. If $2.5 \leq q \leq 4$,

then since $f_{max}/f_{min} = 10^6$, (2-5) yields $\nu = 22$. Thus, a 22-bit phase accumulator is needed, and $N = 2^{22}$. The ROM contains $2^8 = 256$ distinct words, and requires $n + 2 = 10$ input bits. If the frequency resolution and smallest frequency is to be $\Delta \leq 1$ Hz, then (2-2) indicates that $f_r \leq N$ Hz is required so that $f_r = 4$ MHz is satisfactory. When the frequency Δ is desired, the phase increments are so small that $2^{22}/2^{10} = 2^{12}$ increments occur before a new address is specified and a new value of $\sin\theta$ is produced. Thus, the 12 least-significant bits in the accumulator are not used in the addressing of the ROM.□

The direct digital synthesizer can be easily modified to produce a modulated output when high-speed digital data is available. For amplitude modulation, the ROM output is applied to a multiplier. Phase modulation may be implemented by adding the appropriate bits to the phase accumulator output. Frequency modulation entails a modification of the accumulator input bits. For a quaternary modulation, separate sine and cosine ROMs may be used.

The advantages of the digital frequency synthesizer are its ability to produce nearly instantaneous, phase-continuous switching and a very fine frequency resolution and its relatively small size, weight, and power requirements. A disadvantage is the limited maximum frequency, which restricts the bandwidth of the covered frequencies following a frequency translation of the synthesizer output. For this reason, digital frequency synthesizers are often used as components in hybrid synthesizers. Another disadvantage is the stringent requirement for the lowpass filter to suppress frequency spurs generated during changes in the synthesized frequency.

2.3 Indirect Frequency Synthesizer

An *indirect frequency synthesizer* uses voltage-controlled oscillators and feedback loops. The principal components of a single-loop indirect synthesizer, which is similar in operation to a phase-locked loop, are depicted in Figure 8. The control bits, which determine the value of the modulus or divisor N, are supplied by a code generator. The input signal at frequency f_1 may be provided by another synthesizer. Since the feedback loop forces the frequency of the divider output, $(f_0 - f_1)/N$, to closely approximate the reference frequency, f_r, the output of the voltage-controlled oscillator (VCO) is a sine wave with frequency

$$f_0 = Nf_r + f_1 \tag{2-7}$$

where N is a positive integer. Indirect synthesizers usually require less hardware than comparable direct ones, but require more time to switch from one frequency to another. Like digital synthesizers, indirect synthesizers inherently produce phase-continuous outputs after frequency switching.

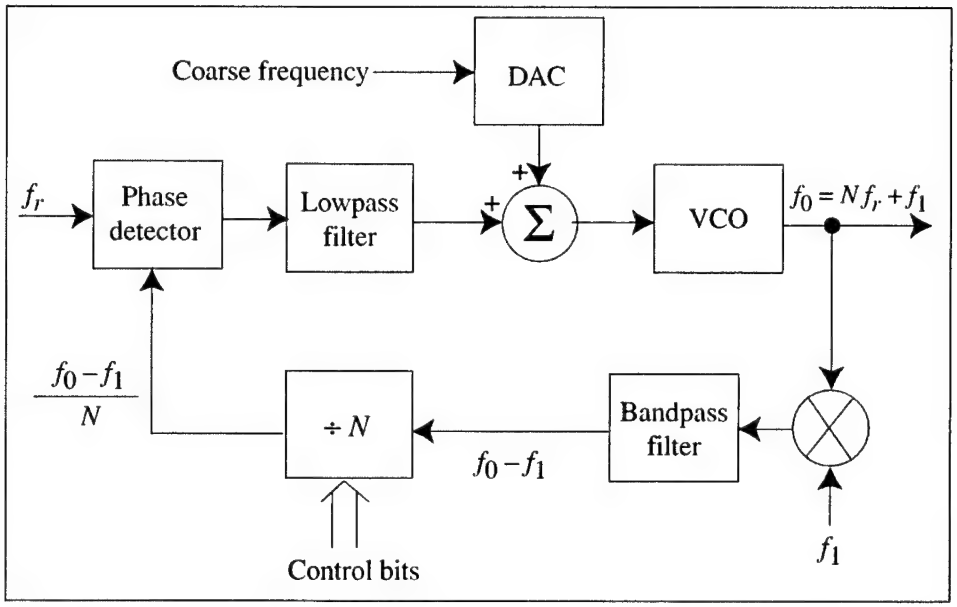


Figure 8. Indirect frequency synthesizer with single loop.

Since the output frequencies change in increments of f_r, the frequency resolution of the single-loop synthesizer is f_r. For stable operation and the suppression of sidebands that are offset from f_0 by f_r, it is desirable that the loop bandwidth be on the order of 0.1 f_r. The *switching time* t_s for changing frequencies, which is inversely proportional to the loop bandwidth, is roughly approximated by

$$t_s = \frac{25}{f_r} \tag{2-8}$$

This equation indicates that a low resolution and a low switching time may not be achievable by a single loop. The switching time t_s is less than or equal to T_{sw} defined previously for frequency-hopping pulses, which may have additional guard time inserted.

Phase detectors in frequency-hopping synthesizers are usually digital devices that measure zero-crossing times rather than the phase differences measured when mixers are used. Digital phase detectors have an extended linear range, are less sensitive to input-level variations, and simplify the interface with a digital divider.

A *divider* is a binary counter that produces a square-wave output. The divider counts down by one unit every time its input crosses zero. If the modulus or divisor is the positive integer N, then after N zero crossings, the divider output crosses zero and changes state. The divider then resumes counting down from N. Programmable dividers have limited operating speeds. Consequently, a divider may not be able to accommodate a high-frequency VCO output. This problem is solved by the down-conversion of the VCO output by the mixer shown in Figure 8, but spurious components are introduced by the mixer. Fixed dividers can operate at much higher speeds than programmable dividers. Therefore, one might

consider placing a fixed divider before the programmable divider in the feedback loop. However, if the fixed divider has a modulus N_1, then the loop resolution becomes $N_1 f_r$ so this solution is usually unsatisfactory.

A *dual-modulus divider*, which is depicted in Figure 9, allows synthesizer operation at high frequencies while maintaining the frequency resolution equal to f_r. The *dual prescalar* consists of two fixed dividers with divisors equal to the positive integers P and $P + Q$. The two programmable dividers count down from the integers A and B, where $B > A$ and A is nonnegative. These dividers are only required to accommodate a frequency much lower than the frequency f_{in} at the input of the dual prescalar. The dual prescalar initially divides by the modulus $P + Q$. This modulus changes whenever a programmable divider reaches zero. After $(P + Q)A$ input transitions, divider 1 reaches zero, and the modulus control causes the dual prescalar to divide by P. Divider 2 has counted down to $B - A$. After $P(B - A)$ more input transitions, divider 2 reaches zero and causes an output transition. The two fixed dividers are then reset, and the dual prescalar reverts to division by $P + Q$. Thus, each output transition corresponds to $A(P + Q) + P(B - A) = AQ + PB$ input transitions, which implies that the dual-modulus divider has a modulus

$$N = AQ + PB\,, \quad B > A \tag{2-9}$$

and produces the output frequency f_{in}/N. If $Q = 1$ and $P = 10$, then the dual-modulus divider is called a *10/11 divider*, and

$$N = 10B + A\,, \quad B > A \tag{2-10}$$

which can be incremented in unit steps by changing A in unit steps.

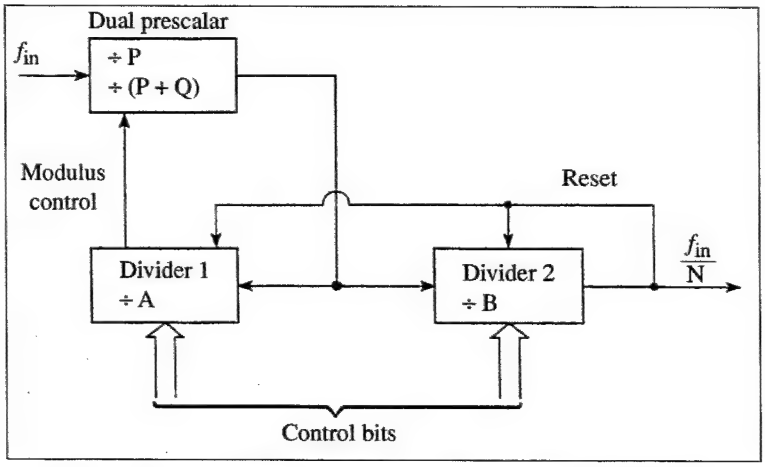


Figure 9. Dual-modulus divider.

Example 3. A frequency-hopping system requires a hopset that covers the VHF band, 30 to 88 MHz, in 25 kHz increments. By taking f_r = 25 kHz, we obtain the desired increment, which is equal to the frequency resolution. Equation (2-8)

indicates that t_s = 1 ms. Equations (2-7) and (2-10) indicate that the band is covered by a 10/11 divider if A ranges from $A_{\min}$ to $A_{\max}$, B ranges from $B_{\min}$ to $B_{\max}$, and

$$f_1 + (10B_{\min} + A_{\min})(25 \cdot 10^3) \leq 30 \cdot 10^6$$
$$f_1 + (10B_{\max} + A_{\max})(25 \cdot 10^3) \geq 88 \cdot 10^6$$

subject to

$$B_{\min} > A_{\max}$$

If we take $A_{\min} = 0$, $A_{\max} = 9$, and $B_{\min} = 10$, then $f_1 \leq 27.5$ MHz and $B_{\max} \geq (351.1 - 4f_1)$ MHz are required. Suitable choices might be $f_1 = 27$ MHz and $B_{\max} = 244$. Thus, A ranges from 0 to 9 while B ranges from 10 to 244. Dividers A and B require 4 and 8 control bits, respectively. If the control bits are stored in a ROM, then each ROM location contains 12 bits. The number of ROM addresses is at least 2321, the number of frequencies in the hopset. Thus, a ROM input address requires 12 bits.□

To decrease the switching time while maintaining the frequency resolution of a single loop, a coarse steering signal can be stored in a ROM, converted into analog form by a digital-to-analog converter (DAC), and applied to the VCO (as shown in Figure 8) immediately after a frequency change. The steering signal reduces the frequency step that must be acquired by the loop when a hop occurs. An alternative approach is to place a fixed divider with modulus M after the loop so that the output frequency is $f_0 = Nf_r/M + f_1/M$. By this means, f_r can be increased without sacrificing resolution provided that the VCO output frequency, which equals Mf_0, is not too large for the divider in the feedback loop.

The switching time can be dramatically reduced by using multiple loops or synthesizers that alternately produce the output frequency. When two synthesizers are used, one synthesizer produces the output frequency while the second one is being tuned to the next frequency following a command from the code generator. If the hop duration exceeds the switching time of each synthesizer, then the second synthesizer begins producing the next frequency before a control switch routes its output to a hopping or dehopping mixer.

A *multiple-loop frequency synthesizer* uses two or more single-loop synthesizers to obtain both fine frequency resolution and fast switching. A three-loop frequency synthesizer is shown in Figure 10. Loops A and B have the form of Figure 8, but loop A does not have a mixer and filter in its feedback. Loop C has the mixer and filter, but lacks the divider. The reference frequency f_r is chosen to ensure that the desired switching time is realized. If $A > M$, then loop C does not appreciably degrade the switching time. The divisor M is chosen so that f_r/M is equal to the desired resolution. Loop A and the divider generate increments of f_r/M while loop B generates increments of f_r. Loop C combines the outputs of loops A and B to produce the output frequency

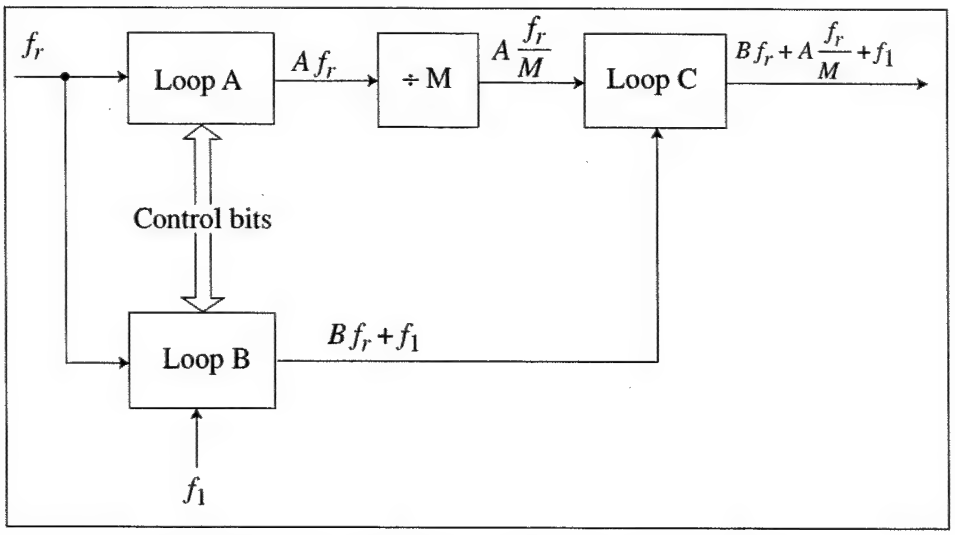


Figure 10. Indirect frequency synthesizer with three loops.

$$f_0 = Bf_r + A\frac{f_r}{M} + f_1 \tag{2-11}$$

where B, A, and M are positive integers because they are produced by dividers. Loop C is preferable to a mixer and bandpass filter because the filter would have to suppress a closely spaced, unwanted component when Af_r/M and Bf_r were far apart. To ensure that each output frequency is produced by unique values of A and B, it is required that

$$A_{\max} = A_{\min} + M - 1 \tag{2-12}$$

According to (2-11), a range of frequencies from $f_{\min}$ to $f_{\max}$ is covered if

$$B_{\min}f_r + A_{\min}\frac{f_r}{M} + f_1 \leq f_{\min} \tag{2-13}$$

$$B_{\max}f_r + A_{\max}\frac{f_r}{M} + f_1 \geq f_{\max} \tag{2-14}$$

To prevent degradation in the switching time, it is required that

$$A_{\min} > M \tag{2-15}$$

Example 4. A hopset must cover 30 to 88 MHz in 25 kHz increments with a t_s = 50 μs. The single-loop synthesizer of Example 3 cannot provide this short switching time. The required switching time is provided by a three-loop synthesizer with f_r = 500 kHz. The resolution of 25 kHz is achieved by taking M = 20. To satisfy (2-15), let $A_{\min}$ = 21. Equation (2-12) then yields $A_{\max}$ = 40. Inequalities (2-13) and (2-14) are satisfied if f_1 = 28 MHz, $B_{\min}$ = 1, and

$B_{max} = 118$. The maximum frequencies that must be accommodated by the dividers in loops A and B are 20 MHz and 59 MHz, respectively. Dividers A and B require 6 and 7 control bits, respectively.□

A *fractional-N synthesizer* uses a single loop and extensive auxiliary hardware to produce an output frequency given by

$$f_0 = \left(N_1 + \frac{i}{M}\right) f_r, \quad i = 0, 1, \ldots, M-1 \tag{2-16}$$

where N_1 and M are positive integers. Although the switching time is inversely proportional to f_r as usual, the resolution is f_r/M, which can be made arbitrarily small in principle. The synthesis method alters the loop feedback by dividing the output frequency by $N_1 + 1$ every M/i output cycles but dividing by N_1 the rest of the time. The effective divisor is then $N = N_1 + i/M$. The implementation entails adding the number i/M to the content of an accumulator every output cycle. Each time the content exceeds unity, a carry bit is generated that causes division by $N_1 + 1$ instead of N_1. For example, if $f_r = 1$ MHz and it is desired to generate $f_0 = 9.15$ MHz, then $N_1 = 9$ and $i/M = 0.15$ is added to the content of an accumulator every output cycle. The output frequency is divided by $N_1 + 1 = 10$ every $M/i = 1/0.15 = 6.67$ output cycles on the average.

The main elements of a fractional-N synthesizer are shown in Figure 11. The *cycle swallower* is a device that blocks one of the VCO output cycles in response to a carry bit from the accumulator. For the VCO to produce a stable output frequency, its input must be approximately a direct-current signal. However, for every reference cycle, the VCO output undergoes N cycles, and the divider output undergoes $N/N_1 = 1 + i/N_1M$ cycles. Therefore, the relative phase between the two phase-detector inputs increases by $2\pi i/N_1M$ radians per

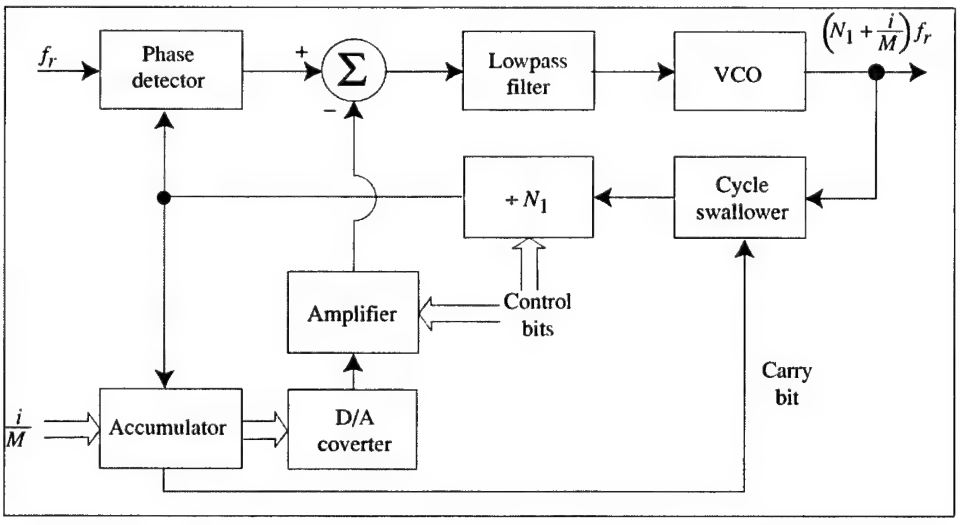


Figure 11. Fractional-N frequency synthesizer.

reference cycle, and the phase-detector output increases at the same rate. This unwanted variation in the phase-detector output is canceled by the amplifier output. Since the accumulator output increases by i/M every reference cycle, a programmable amplifier with a grain of $2\pi/N_1$ yields the output needed for cancellation. The main disadvantage of the fractional-N synthesizer relative to the other synthesizers of comparable performance is its production of relatively high-level spurious signals that frequency-modulate its output signal.

Example 5. A hopset must cover 30 to 88 MHz in 25 kHz increments with t_s = 50 μs. If the output of the fractional-N synthesizer is frequency-translated by 25 MHz, then the synthesizer itself needs to cover 5 MHz to 63 MHz. The switching time is achieved by taking f_r = 500 kHz. The resolution is achieved by taking M = 20. Equation (2-16) indicates that the required frequencies are covered by varying N_1 from 10 to 126 and i from 0 to 19. The accumulator increases its content by $i/M = i/20$ every reference cycle. The integers N_1 and i require 7 and 5 control bits, respectively. □

3. Characteristics of Frequency-Hopping Systems

Frequency hopping may be classified as fast or slow. *Fast frequency hopping* occurs if there is more than one hop for each information symbol. *Slow frequency hopping* occurs if one or more information symbols are transmitted in the time interval between frequency hops. Although these definitions do not refer to the hopping rate, fast frequency hopping is an option only if a hopping rate that exceeds the information-symbol rate can be implemented. Slow frequency hopping is usually preferable because the transmitted waveform is much more spectrally compact (cf. Table 1, Section 5) and the overhead cost of the switching time is reduced.

Let M denote the hopset size, B denote the bandwidth of frequency channels, and F_s denote the minimum separation between adjacent carriers in a hopset. For full protection against stationary narrowband interference and jamming, it is desirable that $F_s \geq B$ so that the frequency channels are nearly spectrally disjoint. A hop then enables the transmitted signal to escape the interference in a frequency channel.

To obtain the full advantage of block or convolutional error-correcting codes in a slow frequency-hopping system, it is important to interleave the code symbols in such a way that the symbol errors in a codeword or constraint length are independent (for hard-decision decoding) or that the symbols are degraded independently (for soft-decision decoding). In frequency-hopping systems operating over a frequency-selective fading channel, the realization of this independence requires certain constraints among the system parameter values. Symbol errors are independent if the fading is independent in each frequency channel and each symbol is transmitted in a different frequency channel. If each of the interleaved code symbols is transmitted at the same location in each hop dwell time, then adjacent symbols are separated by T_h after the interleaving. Thus, a sufficient condition for nearly independent symbol errors is

$$T_h \geq T_{\text{coh}} \tag{3-1}$$

where T_{coh} is the coherence time of the fading channel. Another sufficient condition for nearly independent symbol errors is

$$F_s \geq B_{\text{coh}} \tag{3-2}$$

where B_{coh} is the coherence bandwidth of the fading channel. For practical mobile communication networks with hopping rates exceeding 100 hops/s, (3-1) is rarely satisfied. For a hopping band with bandwidth W and a hopset with a

uniform carrier separation, $F_s = W/M \geq B$. Thus, (3-2) implies that the number of frequency channels is constrained by

$$M \leq \frac{W}{\max(B, B_{\text{coh}})} \tag{3-3}$$

if nearly independent symbol errors are to be ensured. If (3-3) is not satisfied, there will be a performance loss due to the correlated symbol errors. If $B < B_{\text{coh}}$, equalization will not be necessary because the channel transfer function is nearly flat over each frequency channel. If $B \geq B_{\text{coh}}$, either equalization may be used to prevent intersymbol interference or a multicarrier modulation may be combined with the frequency hopping.

Let n denote the length of a block codeword or the constraint length of a convolutional code. Let T_{del} denote the maximum tolerable processing delay. Since the delay caused by coding and ideal interleaving over n hops is $(n-1)T_h + T_s$ and n distinct frequencies are desired,

$$n \leq \min\left(M, 1 + \frac{T_{\text{del}} - T_s}{T_h}\right) \tag{3-4}$$

is required. If this inequality is not satisfied, then nonideal interleaving is necessary, and some performance degradation results.

It is difficult to maintain phase coherence from hop to hop between frequency synthesizers in the transmitter and the receiver, primarily because of frequency-dependent multipath effects and Doppler shifts. Furthermore, the time-varying delay between each received pulse and the synthesizer output in the receiver causes the phase shift in the dehopped signal to differ for each pulse. Thus, practical frequency-hopping systems use noncoherent or differentially coherent demodulators. Unless a pilot signal is available, the hopping rate is very low compared to the transmitted symbol rate, or elaborate iterative demodulation methods are used.

In military applications, the ability of frequency-hopping systems to avoid interference is potentially neutralized by a *repeater jammer* (also known as a *follower jammer*), which is a device that intercepts a signal, processes it, and then transmits jamming at the same center frequency. To be effective against a frequency-hopping system, the jamming energy must reach the victim receiver before it hops to a new set of frequency channels. Thus, the hopping rate is the critical factor in protecting a system against a repeater jammer. Required hopping rates and the limitations of repeater jamming are analyzed in reference [7].

4. Frequency Hopping with MFSK

Multiple frequency-shift keying (MFSK) entails choosing one of q frequencies as the carrier or center frequency for each transmitted symbol in a communication system. When frequency hopping is superimposed on MFSK, resulting in an FH/MFSK system, the set of q possible frequencies changes with each hop. The general transmitter of Figure 2(a) can be simplified for an FH/MFSK system, as illustrated in Figure 12(a), where the code generator output bits and the digital input are combined to determine the frequency generated by the synthesizer.

An FH/MFSK signal has the form

$$s(t) = \sqrt{2S} \sum_{i=-\infty}^{\infty} \sum_{l=0}^{N_h-1} w[t - i(N_h T_s + T_{sw}) - lT_s] \cos[2\pi(f_i + f_{il})t + \phi_i + \phi_{il}] \tag{4-1}$$

where $S = \mathcal{E}_s/T_s$ is the average signal power during a dwell interval, $w(t)$ is a unit-amplitude rectangular pulse of duration T_s, N_h is the number of symbols per

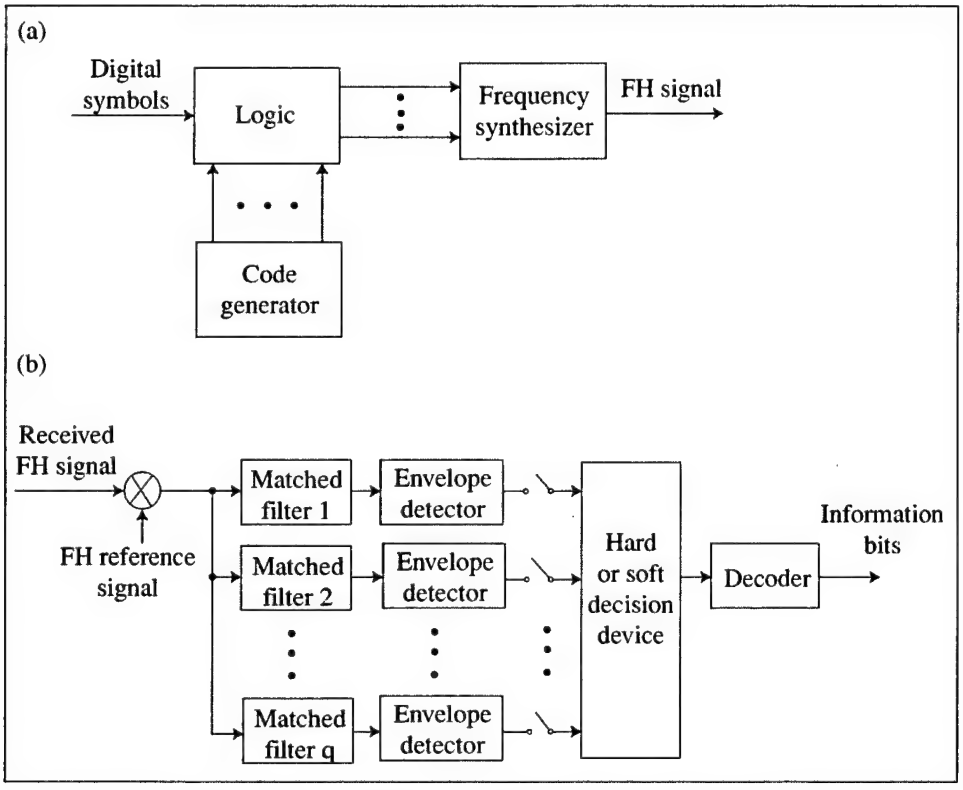


Figure 12. FH/MFSK (a) transmitter and (b) receiver.

dwell interval, f_i is the carrier frequency during dwell interval i, $f_i + f_{il}$ is the MFSK frequency used for symbol l of dwell interval i, ϕ_i is the phase at the beginning of dwell interval i, and ϕ_{il} is the phase associated with MFSK symbol l during dwell interval i. If the MFSK is phase continuous from symbol to symbol, then $\phi_{il} = 0$; otherwise, it may be modeled as a random variable uniformly distributed over [0,2π). The implementation of phase continuity is highly desirable to prevent excessive spectral splatter outside a frequency channel (Section 5).

In an FH/MFSK system, each of the q frequencies or tones in an MFSK set can be considered as the center frequency of an MFSK subchannel. Therefore, the *effective number of frequency channels* is

$$M_e = qM \tag{4-2}$$

where M is the hopset size. In the standard implementation, the q subchannels of each MFSK set are contiguous, and each set constitutes a frequency channel within the hopping band. For noncoherent orthogonal signals, the MFSK tones must be separated enough that a received signal produces negligible responses in the incorrect subchannels. As shown subsequently, the frequency separation must be $f_d = k/T_s$, where k is a nonzero integer, and T_s denotes the symbol duration. To maximize the hopset size when the MFSK subchannels are contiguous, $k = 1$ is selected. Consequently, the bandwidth of a frequency channel for slow frequency hopping with many symbols per dwell interval is

$$B = \frac{q}{T_s} = \frac{q}{T_b \log_2 q} \tag{4-3}$$

where T_b is the duration of a bit, and the factor $\log_2 q$ accounts for the increase in symbol duration when a nonbinary modulation is used. If the hopping band has bandwidth W, the hopset size is

$$M = \left\lfloor \frac{W}{B} \right\rfloor \tag{4-4}$$

where $\lfloor x \rfloor$ denotes the largest integer in x. Figure 12(b) depicts the main elements of a noncoherent FH/MFSK receiver. Each matched filter corresponds to an MFSK subchannel. In practical FH/MFSK systems, the orthogonality of the q MFSK tones is imperfect because of transients that occur after every hop in the receiver.

4.1 Soft-decision decoding

To illustrate some basic issues of frequency-hopping communications and the effectiveness of soft-decision decoding, we consider an FH/MFSK system that uses a repetition code and the receiver of Figure 12(b). Each information symbol, which is transmitted as L code symbols, may be regarded as a codeword or as an

uncoded symbol that uses diversity combining. The interference is modeled as wideband Gaussian noise uniformly distributed over part of the hopping band. Slow frequency hopping with ideal interleaving or fast frequency hopping is assumed to ensure the independence of code-symbol errors. The optimal metric for the Rayleigh-fading channel and a good metric for the additive-white-Gaussian-noise (AWGN) channel without fading is the Rayleigh metric of (A-65), which is

$$U(l) = \sum_{i=1}^{L} R_{li}^2 , \quad l = 1, 2, \ldots , q \tag{4-5}$$

where R_{li} is the sample value of the envelope-detector output that is associated with code symbol i of candidate information-symbol l, and L is the number of repetitions or code symbols. This metric has the advantage that no *side information*, which is specific information about the reliability of symbols, is required for its implementation. A performance analysis of a frequency-hopping system with binary FSK and soft-decision decoding with the *Rayleigh metric* indicates that the system performs poorly against worst-case partial-band jamming [11] primarily because a single jammed frequency can corrupt the metrics. Furthermore, the repetition code is counterproductive because the *noncoherent combining loss* resulting from the fragmentation of the symbol energy is greater than any coding or diversity gain.

Equation (A-64) indicates that a better choice of metric is one with *nonlinear square-law combining* such as

$$U(l) = \sum_{i=1}^{L} \frac{R_{li}^2}{N_{0i}^2} , \quad l = 1, 2, \ldots , q \tag{4-6}$$

where $N_{0i}/2$ is the two-sided power spectral density of the interference and noise over all the MFSK subchannels during code symbol i. A metric that is simpler to implement is the *adaptive gain-control* (AGC) *metric*:

$$U(l) = \sum_{i=1}^{L} \frac{R_{li}^2}{N_{0i}} , \quad l = 1, 2, \ldots , q \tag{4-7}$$

The advantage of both metrics is that they incorporate *side information* contained in the $\{N_{0i}\}$, which are assumed to be known. The subsequent analysis is for the AGC metric, but a slightly more complicated analysis indicates that the metric (4-6) gives exactly the same performance bounds.

The union bound (A-46) implies that the information-symbol error probability satisfies

$$P_{is} \leq (q-1)P_2 \tag{4-8}$$

where P_2 is the probability of an error in comparing the metric associated with the transmitted information symbol with the metric associated with an alternative

one. It is assumed that there are enough frequency channels that L distinct carrier frequencies are used for the L code symbols. Since the MFSK tones are orthogonal, the symbol metrics $\{R_{li}^2/N_{0i}\}$ are independent and identically distributed for all values of l and i (Appendix A-1). Therefore, the Chernoff bound given by (A-107) and (A-106) with $\alpha = 1/2$ yields

$$P_2 \leq \frac{1}{2} Z^L \tag{4-9}$$

$$Z = \min_{0<s<s_1} E\left[\exp\left\{\frac{s}{N_1}\left(R_2^2 - R_1^2\right)\right\}\right] \tag{4-10}$$

where R_1 is the sampled output of an envelope detector when the desired signal is present at the input of the associated matched filter, R_2 is the output when the desired signal is absent, and $N_1/2$ is the two-sided power-spectral density of the interference and noise over all the MFSK subchannels during a code symbol.

For the q-ary symmetric channel, (A-27), (4-8), and (4-9) give an upper bound on the information-bit error probability:

$$P_b \leq \frac{q}{4} Z^L \tag{4-11}$$

For a Gaussian random variable X with mean m and variance σ^2, a direct calculation yields

$$E[\exp(aX^2)] = \frac{1}{\sqrt{1-2a\sigma^2}} \exp\left(\frac{am^2}{1-2a\sigma^2}\right), \quad a < \frac{1}{2\sigma^2} \tag{4-12}$$

From the analysis of Appendix A-1 leading to (A-76), it follows that

$$R_l^2 = x_l^2 + y_l^2, \quad l = 1, 2 \tag{4-13}$$

where x_l and y_l are the real and imaginary parts of R_l, respectively, and are independent Gaussian random variables with the moments

$$E[x_1] = \sqrt{\mathcal{E}_s/2}\cos\theta, \quad E[y_1] = \sqrt{\mathcal{E}_s/2}\sin\theta \tag{4-14}$$

$$E[x_2] = E[y_2] = 0 \tag{4-15}$$

$$\text{var}[x_l] = \text{var}[y_l] = N_1/4, \quad l = 1, 2 \tag{4-16}$$

where $\mathcal{E}_s$ is the energy per symbol. By conditioning on N_1, the expectation in (4-10) can be partially evaluated. Equations (4-12) to (4-16) and the substitution of $\lambda = s/2$ give

$$Z = \min_{0<\lambda<1} E\left[\frac{1}{1-\lambda^2} \exp\left(-\frac{\lambda \mathcal{E}_s/N_1}{1+\lambda}\right)\right] \tag{4-17}$$

where the remaining expectation is over the statistics of N_1.

To simplify the analysis, it is assumed that the thermal noise is negligible. When a repetition symbol encounters no interference, $N_1 = 0$; when it does, $N_1 = I_{t0}/\mu$, where μ is the fraction of the hopping band with interference, and I_{t0} is the spectral density that would exist if the interference power were uniformly spread over the entire hopping band. Since μ is the probability that interference is encountered, (4-17) becomes

$$Z = \min_{0<\lambda<1} \left[\frac{\mu}{1-\lambda^2} \exp\left(-\frac{\lambda\mu\gamma}{1+\lambda} \right) \right] \tag{4-18}$$

where

$$\gamma = \frac{\mathcal{E}_s}{I_{t0}} = \left(\frac{m}{L} \right) \frac{\mathcal{E}_b}{I_{t0}} \tag{4-19}$$

$m = \log_2 q$ is the number of bits per information symbol, and $\mathcal{E}_b$ is the energy per information bit. Using calculus, we find that

$$Z = \frac{\mu}{1-\lambda_0^2} \exp\left(-\frac{\lambda_0\mu\gamma}{1+\lambda_0} \right) \tag{4-20}$$

where

$$\lambda_0 = -\left(\frac{1}{2} + \frac{\mu\gamma}{4} \right) + \left[\left(\frac{1}{2} + \frac{\mu\gamma}{4} \right)^2 + \frac{\mu\gamma}{2} \right]^{1/2} \tag{4-21}$$

Substituting (4-20) and (4-19) into (4-11), we obtain

$$P_b \le 2^{m-2} \left(\frac{\mu}{1-\lambda_0^2} \right)^L \exp\left[-\left(\frac{\lambda_0\mu m}{1+\lambda_0} \right) \frac{\mathcal{E}_b}{I_{t0}} \right] \tag{4-22}$$

which indicates that a repetition code is useless if $\mu = 1$, but if μ is fixed and small, increasing L invariably improves the interference suppression. Thus, the nonlinear diversity combining is potentially effective against stationary interference.

Suppose that the interference is worst-case partial-band jamming. An upper bound on P_b is obtained by maximizing the right-hand side of (4-22) with respect to μ, where $0 \le \mu \le 1$. Calculus yields the maximizing value of μ:

$$\mu_0 = \min\left[\frac{3L}{m} \left(\frac{\mathcal{E}_b}{I_{t0}} \right)^{-1}, \ 1 \right] \tag{4-23}$$

Substituting (4-23) into (4-22), we obtain an upper bound on P_b for worst-case partial-band jamming:

$$P_b \le \begin{cases} 2^{m-2} \left[\dfrac{4L}{me} \left(\dfrac{\mathcal{E}_b}{I_{t0}} \right)^{-1} \right]^L & , \ L \le \dfrac{m}{3} \left(\dfrac{\mathcal{E}_b}{I_{t0}} \right) \\ 2^{m-2}(1-\lambda_0^2)^{-L} \exp\left[-\left(\dfrac{m\lambda_0}{1+\lambda_0} \right) \dfrac{\mathcal{E}_b}{I_{t0}} \right] & , \ L > \dfrac{m}{3} \left(\dfrac{\mathcal{E}_b}{I_{t0}} \right) \end{cases} \tag{4-24}$$

Since μ_0 is obtained by maximizing a bound rather than an equality, it is not necessarily equal to the actual worst-case μ, which would provide a tighter bound than the one in (4-24).

If $\mathcal{E}_b/I_{t0}$ is known, then the number of repetitions can be chosen to minimize the upper bound on P_b for worst-case partial-band jamming. We treat L as a continuous variable such that $L \geq 1$ and let L_0 denote the minimizing value of L. A calculation indicates that the derivative with respect to L of the second line on the right-hand side of (4-24) is positive. Therefore, if $\mathcal{E}_b/I_{t0} < 3/m$, then $L_0 = 1$; if $\mathcal{E}_b/I_{t0} \geq 3/m$, the continuity of (4-24) as a function of L implies that L_0 is determined by the first line in (4-24). Further calculation yields

$$L_0 = \max\left(\frac{m\mathcal{E}_b}{4I_{t0}}\ ,\ 1\right) \tag{4-25}$$

Since L must be an integer, its minimizing value is approximately $\lfloor L_0 \rfloor$. The upper bound on P_b for worst-case partial-band jamming when $L = L_0$ is given by

$$P_b \leq \begin{cases} 2^{m-2}\exp\left(-\dfrac{m\mathcal{E}_b}{4I_{t0}}\right) & ,\quad \dfrac{\mathcal{E}_b}{I_{t0}} \geq \dfrac{4}{m} \\ \dfrac{2^m}{me}\left(\dfrac{\mathcal{E}_b}{I_{t0}}\right)^{-1} & ,\quad \dfrac{3}{m} \leq \dfrac{\mathcal{E}_b}{I_{t0}} < \dfrac{4}{m} \\ 2^{m-2}(1-\lambda_0^2)^{-1}\exp\left[-\left(\dfrac{m\lambda_0}{1+\lambda_0}\right)\dfrac{\mathcal{E}_b}{I_{t0}}\right] & ,\quad \dfrac{\mathcal{E}_b}{I_{t0}} < \dfrac{3}{m} \end{cases} \tag{4-26}$$

This result shows that P_b decreases exponentially as $\mathcal{E}_b/I_{t0}$ increases if the appropriate number of repetitions is chosen and $\mathcal{E}_b/I_{t0}$ is large enough. Thus, the optimal AGC metric sharply limits the performance degradation caused by worst-case partial-band jamming relative to full-band jamming. A comparison of (4-26) and (A-86) with $N_0 \to I_{t0}$ indicates that this degradation is approximately 3 dB for binary FSK. Substituting (4-25) into (4-23), we obtain

$$\mu_0 = \begin{cases} \dfrac{3}{4} & ,\quad \dfrac{\mathcal{E}_b}{I_{t0}} \geq \dfrac{4}{m} \\ \dfrac{3}{m}\left(\dfrac{\mathcal{E}_b}{I_{t0}}\right)^{-1} & ,\quad \dfrac{3}{m} < \dfrac{\mathcal{E}_b}{I_{t0}} < \dfrac{4}{m} \\ 1 & ,\quad \dfrac{\mathcal{E}_b}{I_{t0}} \leq \dfrac{3}{m} \end{cases} \tag{4-27}$$

This result shows that the appropriate choice of L implies that worst-case jamming must cover three-fourths or more of the hopping band, a task that may not be a practical possibility for a jammer.

For frequency hopping with binary FSK and the AGC metric, a more precise derivation [12] that does not use the Chernoff bound and allows $N_0 > 0$ confirms that (4-26) provides an approximate upper bound on the information-bit error rate caused by worst-case partial-band jamming when N_0 is small, although the optimal number of repetitions is much smaller than is indicated by (4-25). Thus,

the appropriate weighting of terms in nonlinear square-law combining prevents the domination by a single corrupted term and limits the inherent noncoherent combining loss.

The implementation of the AGC metric requires the measurement of the interference power. One might attempt to measure this power in frequency channels immediately before the hopping of the signal into those channels, but this method will not be reliable if the interference is frequency-hopping or nonstationary. Another approach is to clip (soft-limit) each envelope-detector output R_{li} to prevent a single erroneous sample from undermining the metric. This method is potentially effective, but its implementation requires an accurate measurement of the signal power for properly setting the clipping level. A sufficiently accurate measurement is often impractical because of fading or power variations across the hopping band. A metric that requires no side information is the *self-normalization metric* defined for binary FSK as [13]

$$U(l) = \sum_{i=1}^{n} \frac{R_{li}^2}{R_{1i}^2 + R_{2i}^2}, \quad l = 1, 2 \tag{4-28}$$

Although it does not provide as good a performance against partial-band jamming as the ideal AGC metric, the self-normalization metric is far more practical and is generally superior to hard-decision decoding.

It has been assumed that all subchannels in an MFSK set are jammed or none are. However, this assumption ignores the threat of narrowband jamming signals that are randomly distributed over the frequency channels. Although (4-26) indicates that it is advantageous to use nonbinary signaling when $\mathcal{E}_b/I_{t0} \geq 4/m$, this advantage is completely undermined when distributed, narrowband jamming signals are a threat. A fundamental problem, which also limits the applicability of FH/MFSK in networks, is the reduced hopset size for nonbinary MFSK indicated by (4-4) and (4-3).

4.2 Narrowband jamming signals

When the MFSK subchannels are contiguous, it is not advantageous to a jammer to transmit the jamming in all the subchannels of an MFSK set because only a single subchannel needs to be jammed to cause a symbol error. A sophisticated jammer with knowledge of the spectral locations of the MFSK sets can cause increased system degradation by placing one jamming tone or narrowband jamming signal in every MFSK set.

To assess the impact of this sophisticated multitone jamming on the receiver of Figure 12(b), it is assumed that thermal noise is absent and that each jamming tone coincides with one MFSK tone in a frequency channel encompassing q MFSK tones [8], [9]. Whether a jamming tone coincides with the transmitted MFSK tone or an incorrect one, there will be no symbol error if the desired-signal power S exceeds the jamming power. Thus, if I_t is the total available

jamming power, then the jammer can maximize symbol errors by placing tones with power levels slightly above S whenever possible in approximately J frequency channels such that

$$J = \begin{cases} 1 & , \quad I_t < S \\ \left\lfloor \frac{I_t}{S} \right\rfloor & , \quad S \le I_t \le MS \\ M & , \quad MS < I_t \end{cases} \tag{4-29}$$

If a transmitted tone enters a jammed frequency channel and $I_t \ge S$, then with probability $(q-1)/q$ the jamming tone will not coincide with the transmitted tone and will cause a symbol error. If the jamming tone does coincide with the correct tone, it may cause a symbol error in the absence of thermal noise only if its power level is exactly S and it has exactly a 180° phase shift relative to the desired signal, an event with zero probability. Since J/M is the probability that a frequency channel is jammed, and no error occurs if $I_t < S$, the symbol error probability is

$$P_s = \begin{cases} 0 & , \quad I_t < S \\ \frac{J}{M}\left(\frac{q-1}{q}\right) & , \quad I_t \ge S \end{cases} \tag{4-30}$$

Substitution of (4-3), (4-4), and (4-29) into (4-30) and the approximation $\lfloor x \rfloor \approx x$ yields

$$P_s = \begin{cases} \frac{q-1}{q} & , \quad \frac{\mathcal{E}_b}{I_{t0}} < \frac{q}{\log_2 q} \\ \left(\frac{q-1}{\log_2 q}\right)\left(\frac{\mathcal{E}_b}{I_{t0}}\right)^{-1} & , \quad \frac{q}{\log_2 q} \le \frac{\mathcal{E}_b}{I_{t0}} \le WT_b \\ 0 & , \quad \frac{\mathcal{E}_b}{I_{t0}} > WT_b \end{cases} \tag{4-31}$$

where $\mathcal{E}_b = ST_b$ denotes the energy per bit and $I_{t0} = I_t/W$ denotes the spectral density of the interference power that would exist if it were uniformly spread over the hopping band. This equation exhibits an inverse linear dependence of P_s on $\mathcal{E}_b/I_{t0}$, which indicates that the jamming has an impact qualitatively similar to that of Rayleigh fading. It is observed that P_s increases with q, which is the opposite of what is observed over the AWGN channel. Thus, binary FSK is advantageous against this sophisticated multitone jamming.

To preclude this jamming, each MFSK tone in an MFSK set may be independently hopped. However, this approach demands a large increase in the amount of hardware, and uniformly distributed, narrowband jamming signals are almost as damaging as the worst-case multitone jamming. Thus, contiguous MFSK subchannels are usually preferable, and the FH/MFSK receiver has the form of Figure 12(b). An analysis of FH/MFSK systems with hard-decision

decoding in the presence of uniformly distributed, narrowband jamming signals confirms the superior robustness of binary FSK relative to nonbinary MFSK whether the MFSK tones hop independently or not [10].

5. Modulations

In a network of frequency-hopping systems, it is highly desirable to choose a spectrally compact modulation so that the number of frequency channels is large and, hence, the number of collisions between frequency-hopping signals is kept small. As discussed in Section 4, binary orthogonal FSK allows more frequency channels than MFSK and, hence, is advantageous against narrowband interference distributed throughout the hopping band. A spectrally compact modulation helps ensure that $B < B_{coh}$ so that equalization in the receiver is not necessary, as discussed in Section 3. This section considers spectrally compact alternatives to orthogonal FSK.

The demodulator transfer function following the dehopping in Figure 2 is assumed to have a bandwidth approximately equal to B, the bandwidth of a frequency channel. The bandwidth is determined primarily by the percentage of the signal power that must be processed by the demodulator if the demodulated signal distortion and the intersymbol interference are to be negligible. In practice, this percentage must be at least 90 percent and is often more than 95 percent. The relation between B and the symbol duration may be expressed as

$$B = \frac{\zeta}{T_s} \tag{5-1}$$

where ζ is a constant determined by the signal modulation. For example, if minimum-shift keying is used, the transfer function is rectangular, and many symbols are transmitted during a dwell interval, then $\zeta = 0.8$ if 90 percent of the signal power is included in a frequency channel, and $\zeta = 1.2$ if 99 percent is included.

Spectral splatter is the interference produced in frequency channels other than the one being used by a frequency-hopping pulse. It is caused by the time-limited nature of transmitted pulses. The degree to which spectral splatter may cause errors depends primarily on F_s (see Section 3) and the percentage of the signal power included in a frequency channel. Usually, only pulses in adjacent channels produce a significant amount of spectral splatter in a frequency channel.

The *adjacent splatter ratio* K_s is the ratio of the power due to spectral splatter from an adjacent channel to the corresponding power that arrives at the receiver in that channel. For example, if B is the bandwidth of a frequency channel that includes 97 percent of the signal power and $F_s \geq B$, then no more than 1.5 percent of the power from a transmitted pulse can enter an adjacent channel on one side of the frequency channel used by the pulse; therefore, $K_s \leq 0.015$. A given maximum value of K_s can be reduced by an increase in F_s, but eventually

the value of M must be reduced if W is fixed. As a result, the rate at which users hop into the same channel increases. This increase may cancel any improvement due to the reduction of the spectral splatter. The opposite procedure (reducing F_s and B so that more frequency channels become available) increases not only the spectral splatter but also signal distortion and intersymbol interference, so the amount of useful reduction is limited.

To avoid spectral spreading due to amplifier nonlinearities, it is desirable for the signal modulation to have a constant envelope, as it is often impossible to implement a filter with the appropriate bandwidth and center frequency for spectral shaping of a signal after it emerges from the final power amplifier. Noncoherent demodulation is nearly always a practical necessity in frequency-hopping systems unless the dwell interval is large. Accordingly, good modulation candidates are differential phase-shift keying (DPSK) and minimum-shifting keying (MSK) or some other form of spectrally compact continuous-phase frequency-shift keying (CPFSK) or continuous-phase modulation (CPM).

The general form of a CPM signal is

$$s(t) = A\cos[2\pi f_c t + \phi(t, \boldsymbol{\alpha})] \tag{5-2}$$

where A is the amplitude, f_c is the carrier frequency, and $\phi(t, \boldsymbol{\alpha})$ is the phase function that carries the message. The phase function has the ideal form

$$\phi(t, \boldsymbol{\alpha}) = 2\pi h \int_{-\infty}^{t} \left[\sum_{i=-\infty}^{\infty} \alpha_i g(x - iT_s) \right] dx \tag{5-3}$$

where h is a constant called the *deviation ratio* or *modulation index*, T_s is the symbol duration, and the vector $\boldsymbol{\alpha}$ is a sequence of q-ary channel symbols. Each symbol α_i takes one of q values; if q is even,

$$\alpha_i = \pm 1, \pm 3, \ldots, \pm(q-1), \quad i = 0, 1, 2, \ldots \tag{5-4}$$

Equation (5-3) exhibits the phase continuity and indicates that the phase in any specified symbol interval depends on the previous symbols.

It is assumed that the integrand in (5-3) is piecewise continuous so that $\phi(t, \boldsymbol{\alpha})$ is differentiable. The frequency function of the CPM signal, which is proportional to the derivative of $\phi(t, \boldsymbol{\alpha})$, is

$$\frac{1}{2\pi}\phi'(t, \boldsymbol{\alpha}) = h \sum_{i=-\infty}^{\infty} \alpha_i g(t - iT_s) \tag{5-5}$$

The frequency pulse $g(t)$ is assumed to vanish outside an interval; that is,

$$g(t) = 0\,, \quad t < 0\,, \quad t > LT_s \tag{5-6}$$

where L is a positive integer and may be infinite. The presence of h as a multiplicative factor in the pulse function makes it convenient to normalize $g(t)$ by assuming that

$$\int_0^{LT_s} g(x)dx = \frac{1}{2} \tag{5-7}$$

If $L = 1$, the continuous phase modulation is called a *full-response modulation*; if $L > 1$, it is called a *partial-response modulation*, and each frequency pulse extends over two or more symbol intervals. The normalization condition for a full-response modulation implies that the phase change over a symbol interval is equal to $h\pi\alpha_i$.

Continuous-phase frequency-shift keying (CPFSK) is a subclass of CPM for which the instantaneous frequency is constant over each symbol interval. Because of the normalization, a CPFSK frequency pulse is given by

$$g(t) = \begin{cases} \dfrac{1}{2LT_s}, & 0 \le t \le LT_s \\ 0, & \text{otherwise} \end{cases} \tag{5-8}$$

A binary CPFSK signal shifts between two frequencies separated by $f_d = h/T_s$. *Minimum-shift keying* is defined as binary CPFSK with h = 1/2 and, hence, frequencies are separated by $f_d = 1/2T_s$. The main difference between CPFSK and MFSK is that h can have any positive value for CPFSK but is relegated to integer values for MFSK so that the tones are orthogonal to each other. A second difference is that MFSK is detected with matched filters and envelope detectors, whereas CPFSK with $h < 1$ is usually detected with a frequency discriminator. Although CPFSK explicitly requires phase continuity and MFSK does not, MFSK is usually implemented with phase continuity to avoid the generation of spectral splatter.

A measure of the spectral compactness of signals is provided by the *fractional out-of-band* power defined as

$$P_{ob}(f) = 1 - \frac{\int_{-f}^{f} S(f')df'}{\int_{-\infty}^{\infty} S(f')df'}, \quad f \ge 0 \tag{5-9}$$

where f is the frequency variable and $S(f)$ is the two-sided power spectral density of the complex envelope of the signal (Appendix B-3), which is often called the *equivalent lowpass waveform*. The closed-form expressions for the power spectral densities of QPSK and binary MSK (Appendix B) can be used to generate Figure 13. The curves depict $P_{ob}(f)$ in decibels as a function of f in units of $1/T_b$, where $T_b = T_s/\log_2 q$ for a q-ary modulation. The fractional power within a transmission channel of one-sided bandwidth B is given by

$$K_0 = 1 - P_{ob}(B/2) \tag{5-10}$$

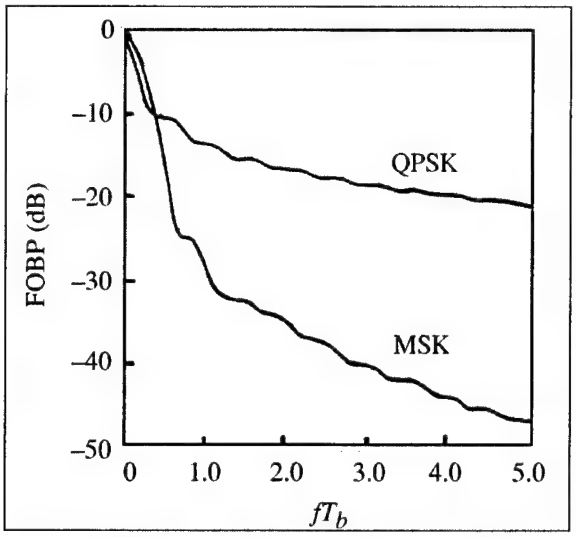


Figure 13. Fractional out-of-band power (FOBP) for equivalent lowpass waveforms of QPSK and MSK.

Usually, K_0 must exceed at least 0.9 to prevent significant performance degradation in communications over a bandlimited channel. The transmission bandwidth for which $K_0 = 0.99$ is approximately $1.2/T_b$ for binary MSK, but approximately $8/T_b$ for PSK or QPSK. The adjacent splatter ratio, which is due to out-of-band power on one side of the center frequency, has the upper bound given by

$$K_s < \frac{1}{2} P_{ob}(B/2) \tag{5-11}$$

An FH/CPM signal has a continuous phase over each dwell interval with N symbols but has a phase discontinuity every $T_h = NT_s + T_{sw}$ seconds at the beginning of another dwell interval. The signal may be expressed as

$$s(t) = \sqrt{2S} \sum_{i=-\infty}^{\infty} w(t - iT_h, T_d) \cos\left[2\pi f_i t + \phi(t, \boldsymbol{\alpha}) + \theta_i\right] \tag{5-12}$$

where $S = \mathcal{E}_s/T_s$ is the average signal power during a dwell interval, $w(t, T_d)$ is a unit-amplitude rectangular pulse of duration $T_d = NT_s$, f_i is the carrier frequency during hop-interval i, and θ_i is the phase at the beginning of dwell-interval i.

Consider multitone jamming of an FH/CPM or FH/CPFSK system in which the thermal noise is absent and each jamming tone is randomly placed within a single frequency channel. It is reasonable to assume that a symbol error occurs with probability $(q-1)/q$ when the frequency channel contains a jamming tone with power exceeding S. Since (4-3) is not applicable to CPM or CPFSK, a derivation similar to that of (4-31) yields

$$P_s = \begin{cases} \dfrac{q-1}{q} & , \quad \dfrac{\mathcal{E}_b}{I_{t0}} < BT_b \\ \left(\dfrac{q-1}{q}\right) BT_b \left(\dfrac{\mathcal{E}_b}{I_{t0}}\right)^{-1} & , \quad BT_b \le \dfrac{\mathcal{E}_b}{I_{t0}} \le WT_b \\ 0 & , \quad \dfrac{\mathcal{E}_b}{I_{t0}} > WT_b \end{cases} \tag{5-13}$$

for sophisticated multitone jamming, where T_b is the bit duration. Since the orthogonality of the MFSK tones is not a requirement for CPM or CPFSK, the bandwidth B for FH/CPM or FH/CPFSK may be much smaller than the bandwidth for FH/MFSK.

Consider multitone jamming of an FH/DPSK system with negligible thermal noise. Each tone is assumed to have a frequency identical to the center frequency of one of the frequency channels. A DPSK demodulator compares the phases of two successive received symbols. If the magnitude of the phase difference is less then $\pi/2$, then the demodulator decides that a 1 was transmitted; otherwise, it decides that a 0 was transmitted. The composite signal, consisting of the transmitted signal plus the jamming tone, has a constant phase over two successive received symbols in the same dwell interval, if a 1 was transmitted and the thermal noise is absent; thus, the demodulator will correctly detect the 1.

Suppose that a 0 was transmitted. Then the desired signal is $\sqrt{2S}\cos 2\pi f_c t$ during the first symbol and $-\sqrt{2S}\cos 2\pi f_c t$ during the second symbol, respectively, where f_c is the carrier frequency of the frequency-hopping signal during the dwell interval. When a jamming tone is present, trigonometric identities indicate that the composite signal during the first symbol may be expressed as

$$\sqrt{2S}\cos 2\pi f_c t + \sqrt{2I}\cos\left(2\pi_c t + \theta\right) = \sqrt{2S + 2I_t + 4\sqrt{SI}\cos\theta}\cos\left(2\pi f_c t + \phi_1\right) \tag{5-14}$$

where I is the average power of the tone, θ is the phase of the tone relative to the phase of the transmitted signal, and ϕ_1 is the phase of the composite signal:

$$\phi_1 = \tan^{-1}\left(\frac{\sqrt{I}\sin\theta}{\sqrt{S} + \sqrt{I}\cos\theta}\right) \tag{5-15}$$

Since the desired signal during the second symbol is $-\sqrt{2S}\cos 2\pi f_c t$, the phase of the composite signal during the second symbol is

$$\phi_2 = \tan^{-1}\left(\frac{\sqrt{I}\sin\theta}{-\sqrt{S} + \sqrt{I}\cos\theta}\right) \tag{5-16}$$

Using trigonometry, it is found that

$$\cos(\phi_2 - \phi_1) = \frac{I - S}{\sqrt{S^2 + I^2 + 2SI(1 - 2\cos^2\theta)}} \tag{5-17}$$

If $I > S$, $|\phi_2 - \phi_1| < \pi/2$ so the demodulator incorrectly decides that a 1 was transmitted. If $I < S$, no mistake is made. Thus, multitone jamming with total power I_t is most damaging when J frequency channels given by (4-29) are jammed and each tone has power $I = I_t/J$. If the information bits 0 and 1 are equally likely, then the symbol error probability given that a frequency channel is jammed with $I > S$ is P_s = 1/2, the probability that a 0 was transmitted. Therefore, $P_s = J/2M$ if $I_t \geq S$, and $P_s = 0$, otherwise. Using (4-4) and (4-29) with $S = E_b/T_b$, $I_t = I_{t0}W$, and $\lfloor x \rfloor \approx x$, we obtain the symbol error probability for DPSK and multitone jamming:

$$P_s = \begin{cases} \frac{1}{2}, & \frac{\mathcal{E}_b}{I_{t0}} < BT_b \\ BT_b\left(\frac{\mathcal{E}_b}{I_{t0}}\right)^{-1}, & BT_b \leq \frac{\mathcal{E}_b}{I_{t0}} \leq WT_b \\ 0, & \frac{\mathcal{E}_b}{I_{t0}} > WT_b \end{cases} \tag{5-18}$$

The same result holds for binary CPFSK.

As implied by Figure 13, the bandwidth requirement of DPSK with $K_0 > 0.9$, which is the same as that of PSK or QPSK and less than that of orthogonal FSK, exceeds that of MSK. Thus, if the hopping bandwidth W is fixed, the number of frequency channels available for FH/DPSK is smaller than it is for noncoherent FH/MSK. This increase in B and reduction in frequency channels offsets the intrinsic performance advantage of DPSK and implies that noncoherent FH/MSK will give a lower P_s than FH/DPSK in the presence of worst-case multitone jamming, as indicated in (5-18). Alternatively, if the bandwidth of a frequency channel is fixed, an FH/DPSK signal will experience more distortion and spectral splatter than an FH/MSK signal. Any pulse shaping of the DPSK symbols will alter their constant envelope. An FH/DPSK system is more sensitive to Doppler shifts and frequency instabilities than an FH/MSK system. Another disadvantage of FH/DPSK is due to the usual lack of phase coherence from hop to hop, which necessitates an extra phase-reference symbol at the start of every dwell interval. This extra symbol reduces $\mathcal{E}_s$ by a factor $(N_h - 1)/N_h$, where N_h is the number of symbols per hop or dwell interval and $N_h \geq 2$. Thus, DPSK does not appear to be as suitable a means of modulation as noncoherent MSK for most applications of frequency-hopping communications, and the main competition for MSK comes from other forms of CPM.

The *cross-correlation parameter* for two signals $s_1(t)$ and $s_2(t)$, each with energy $\mathcal{E}_s$, is defined as

$$C = \frac{1}{\mathcal{E}_s}\int_0^{T_s} s_1(t)s_2(t)dt \tag{5-19}$$

For CPFSK, two possible transmitted signals, each representing a different channel symbol, are

$$s_1(t) = \sqrt{2\mathcal{E}_s/T_s}\cos(2\pi f_1 t + \phi_1)\ , \quad s_2(t) = \sqrt{2\mathcal{E}_s/T_s}\cos(2\pi f_2 t + \phi_2) \tag{5-20}$$

The substitution of these equations into (5-19), a trigonometric expansion and discarding of an integral that is negligible if $(f_1 + f_2)T_d >> 1$, and the evaluation of the remaining integral give

$$C = \frac{1}{2\pi f_d T_s}[\sin(2\pi f_d T_s + \phi_d) - \sin\phi_d]\ , \quad f_d \neq 0 \tag{5-21}$$

where $f_d = f_1 - f_2$ and $\phi_d = \phi_1 - \phi_2$. Because of the phase synchronization in a coherent demodulator, we may take $\phi_d = 0$. Therefore, the orthogonality condition $C = 0$ is satisfied if $h = f_d T_s = k/2$, where k is any nonzero integer. The smallest value of h for which $C = 0$ is $h = 1/2$, which corresponds to MSK.

In a noncoherent demodulator, ϕ_d is a random variable, which is assumed to be uniformly distributed over [0, 2π). Equation (5-21) indicates that $E[C] = 0$ for all values of h. The variance of C is

$$\begin{aligned}\text{var(C)} &= \left(\frac{1}{2\pi f_d T_s}\right)^2 E\left[\sin^2(2\pi f_d T_s + \phi_d) + \sin^2\phi_d - 2\sin\phi_d \sin(2\pi f_d + \phi_d)\right] \\ &= \left(\frac{1}{2\pi f_d T_s}\right)^2 (1 - \cos 2\pi f_d T_s) \\ &= \frac{1}{2}\left(\frac{\sin\pi h}{\pi h}\right)^2 \end{aligned} \tag{5-22}$$

Since var(C) $\neq 0$ for $h = 1/2$, MSK does not provide orthogonal signals for noncoherent demodulation. If h is any nonzero integer, then both (5-22) and (5-21) indicate that the two CPFSK signals are orthogonal for any ϕ_d. This result justifies the previous assertion that MFSK tones must be separated by $f_d = k/T_s$ to provide noncoherent orthogonal signals.

A noncoherent FH/CPFSK signal can be represented by (5-12). The power spectral density of the complex envelope of this signal, which is the same as the dehopped power spectral density, depends on the number of symbols per dwell interval, N_h, because of the random phases $\{\theta_i\}$. The power spectral density has been calculated [14] for binary CPFSK, assuming that each θ_i is an independent random variable uniformly distributed over [0,2π) and the information symbols are ± 1 with equal probability. The 99 percent bandwidths of FH/CPFSK with deviation ratios $h = 0.5$ and $h = 0.7$ are listed in Table 1 for different values of N_h. As N_h increases, the power spectral density becomes more compact and approaches that of coherent CPFSK without frequency hopping. For $N_h \geq 64$, the frequency hopping causes little spectral spreading. However, fast frequency hopping, which corresponds to $N_h = 1$, entails a very large 99 percent bandwidth.

Table 1. 99 percent bandwidth in $1/T_s$ versus symbols per dwell interval for FH/CPFSK.

	Deviation ratio	
Symbols/dwell	$h = 0.5$	$h = 0.7$
1	18.844	18.688
2	9.9375	9.9688
4	5.1875	5.2656
16	1.8906	2.1250
64	1.2813	1.8750
256	1.2031	1.8125
1024	1.1875	1.7969
No hopping	1.1875	1.7813

This fact is the main reason why slow frequency hopping is usually preferable to fast frequency hopping.

With multisymbol noncoherent detection, full-response CPFSK systems can provide a better symbol error probability than coherent PSK systems [15]. For r-symbol detection, where r is odd, the optimal receiver correlates the received waveform over all possible r-symbol patterns before making a decision about the middle symbol. The drawback is the considerable implementation complexity of multisymbol detection, even for three-symbol detection. An additional problem for FH/CPFSK with multisymbol detection is that the first and last $(r-1)/2$ symbols during a dwell interval cannot use the multisymbol detection without accessing other dwell intervals, which may cause practical difficulties.

Symbol-by-symbol noncoherent detection after the dehopping of the FH/CPFSK signal can be inexpensively implemented by using a limiter and frequency discriminator, as illustrated in Figure 14. Analysis of the *limiter-discriminator* or *frequency discriminator* [16] provides complicated expressions for the symbol error probability in the presence of white Gaussian noise. However, the theoretical P_s can be approximated to within a few tenths of a decibel by

$$P_s = \frac{1}{2}\exp\left(-\xi\frac{\mathcal{E}_s}{N_0}\right) \tag{5-23}$$

where the parameter ξ depends on h and the product BT_s, and $N_0/2$ is the two-sided power spectral density of the noise. If the frequency discriminator has a Gaussian filter and $BT_s = 1$, then it is found that P_s is minimized when $h \approx 0.7$. For CPFSK with $h = 0.7$ and $BT_s = 1$, setting $\xi = 0.7$ in (5-23) provides an approximate least-squares fit to the theoretical curve for P_s over the range

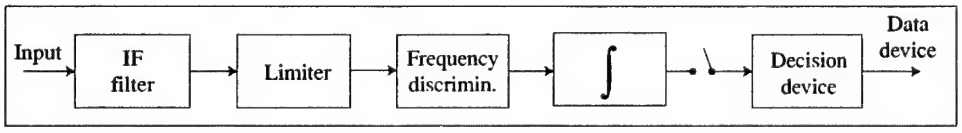


Figure 14. Frequency discriminator for CPFSK.

$10^{-6} \leq P_s \leq 10^{-2}$. If $BT_s = 1$, then $\xi = 0.5$ provides a close fit over the same range for orthogonal CPFSK with $h = 1$ and a fairly close fit for MSK($h = 0.5$). Thus, the discriminator demodulation of MSK or orthogonal CPFSK provides approximately the same performance as envelope detection of orthogonal FSK. The practical difference is that noncoherent MSK requires roughly half the bandwidth of orthogonal FSK for acceptable levels of spectral splatter and intersymbol interference. The increased number of frequency channels due to the decreased value of B does not give FH/MSK an advantage over the AWGN channel. However, the increase is advantageous against a fixed number of interference tones, optimized jamming, and multiple-access interference in a network of frequency-hopping systems, as discussed in the next section.

Since $\xi = 0.7$ for an FH/CPFSK system with $h = 0.7$, this system has a potential 1.46 dB advantage in $\mathcal{E}_s$ relative to an FH/MSK system with $BT_s = 1$. However, since CPFSK with $h = 0.7$ does not have as compact a spectrum as MSK, the FH/CPFSK system will have increased intersymbol interference due to bandlimiting and spectral splatter relative to the FH/MSK system. Only if these effects are negligible can the potential 1.46 dB advantage be realized. When $N_s \geq 64$, reducing the spectral splatter of the FH/CPFSK to the same level it is for FH/MSK with $B = 1/T_s$ requires that $B = 1.4/T_s$. The increased bandwidth lowers ξ and decreases the number of frequency channels.

6. Error-Correcting Codes for Partial-Band Interference

When partial-band interference is present, let I_{t0} denote the one-sided interference power spectral density that would exist if the power were uniformly distributed over the hopping band. If a fixed amount of interference power is uniformly distributed over J frequency channels out of M in the hopping band, then the fraction of the hopping band with interference is

$$\mu = \frac{J}{M} \tag{6-1}$$

and the interference power spectral density in each of the interfered channels is I_{t0}/μ. When the frequency-hopping signal uses a carrier frequency that lies within the spectral region occupied by the partial-band interference, this interference is modeled as additional white Gaussian noise that increases the noise-power spectral density from N_0 to $N_0 + I_{t0}/\mu$. Therefore, for hard-decision decoding, the symbol error probability is

$$P_s = \mu F\left(\frac{\mathcal{E}_s}{N_0 + I_{t0}/\mu}\right) + (1-\mu)F\left(\frac{\mathcal{E}_s}{N_0}\right) \tag{6-2}$$

where the conditional symbol error probability $F(\ )$ depends on the modulation and fading. For noncoherent FH/MFSK and Ricean fading, (A-89) implies that

$$F(x) = \sum_{i=1}^{q-1}(-1)^{i+1}\binom{q-1}{i}\frac{\kappa+1}{\kappa+1+(\kappa+1+x)i}\exp\left[-\frac{\kappa x i}{\kappa+1+(\kappa+1+x)i}\right] \tag{6-3}$$

where q is the alphabet size of the MFSK symbols and κ is the Rice factor. When there is no fading and the modulation is binary CPFSK, then (5-23) implies that

$$F(x) = \frac{1}{2}\exp(-\xi x) \tag{6-4}$$

For the AWGN channel and no fading, classical communication theory indicates that $F(x)$ for DPSK is given by (6-4) with $\xi = 1$. However, $\mathcal{E}_s$ in (6-2) must be reduced by the factor $N_h/(N_h+1)$ because of the reference symbol that must be included in each dwell interval.

If μ is treated as a continuous variable over [0,1] and $I_{t0} >> N_0$, then straightforward calculations using (6-2) and (6-4) indicate that the worst-case value of μ is

$$\mu_0 = \min\left[\left(\frac{\xi\mathcal{E}_s}{I_{t0}}\right)^{-1},\ 1\right] \tag{6-5}$$

The corresponding worst-case symbol error probability is

$$P_s = \begin{cases} \dfrac{1}{2e}\left(\dfrac{\xi\mathcal{E}_s}{I_{t0}}\right)^{-1}, & \dfrac{\xi\mathcal{E}_s}{I_{t0}} \geq 1 \\ \dfrac{1}{2e}\exp\left(-\dfrac{\xi\mathcal{E}_s}{I_{t0}}\right), & \dfrac{\xi\mathcal{E}_s}{I_{t0}} < 1 \end{cases} \tag{6-6}$$

which does not depend on M because of the assumption that μ is a continuous variable. For Rayleigh fading and binary CPFSK, similar calculations using (6-3) with $q = 2$ and $\kappa = 0$ yield $\mu_0 = 1$. Thus, in the presence of Rayleigh fading, interference spread uniformly over the entire hopping band hinders communications more than interference concentrated over part of the band.

For FH/MFSK with an error-correcting code, the bandwidth of a frequency channel must be increased to $B = qB_u/2(\log_2 q)r$, where $r = k/n$ is the code rate and B_u is the bandwidth for binary FSK in the absence of coding. If W, the bandwidth of the hopping band, is fixed, then the number of disjoint frequency channels available for hopping is reduced to

$$M = \left\lfloor \frac{2(\log_2 q)rW}{qB_u} \right\rfloor \tag{6-7}$$

The energy per channel symbol is

$$\mathcal{E}_s = r(\log_2 q)\mathcal{E}_b \tag{6-8}$$

When the interference is partial-band jamming, J and, hence, μ are parameters that may be varied by a jammer. It is assumed henceforth that M is large enough that μ in (6-2) may be treated as a continuous variable over $[1/M, 1]$. With this assumption, the error probabilities do not explicitly depend on M.

6.1 Reed-Solomon Codes

The use of a Reed-Solomon code with MFSK is advantageous against partial-band interference for two principal reasons. First, a Reed-Solomon code is maximum-distance-separable (Appendix A-1) and hence accommodates many erasures. Second, the use of nonbinary MFSK symbols to represent code symbols allows a relatively large symbol energy, as indicated by (6-8).

Consider an FH/MFSK system that uses a Reed-Solomon code with no erasures in the presence of partial-band interference and Ricean fading. The demodulator comprises a parallel bank of noncoherent detectors and a device that makes hard decisions. In a slow frequency-hopping system, symbol interleaving among different dwell intervals and subsequent deinterleaving in the receiver may be needed to disperse errors due to the fading or interference and thereby facilitate their removal by the decoder. In a fast frequency-hopping system, symbol errors may be independent so that interleaving is unnecessary. The MFSK modulation

implies a q-ary symmetric channel. Therefore, for ideal symbol interleaving and hard-decision decoding of loosely packed codes, (A-26) and (A-27) indicate that

$$P_b \approx \frac{q}{2(q-1)} \sum_{i=t+1}^{n} \binom{n-1}{i-1} P_s^i (1-P_s)^{n-i} \tag{6-9}$$

Figure 15 shows P_b for FH/MFSK and an extended Reed-Solomon (32,12) code in the presence of Ricean fading. The frequency channels are assumed to be separated enough that fading events are independent. Thus, (6-2), (6-3), and (6-9) are applicable. For $\kappa > 0$, the curves exhibit peaks as the fraction of the band with interference varies. These peaks indicate that for a specific value of $\mathcal{E}_b/I_{t0}$, the concentration of the interference power over part of the hopping band (perhaps intentionally by a jammer) is more damaging than uniformly distributed interference. The peaks become sharper and occur at smaller values of μ as $\mathcal{E}_b/I_{t0}$ increases. For Rayleigh fading, which corresponds to $\kappa = 0$, peaks are absent in the figure, and full-band interference is the most damaging. As κ increases, the peaks appear and become more pronounced.

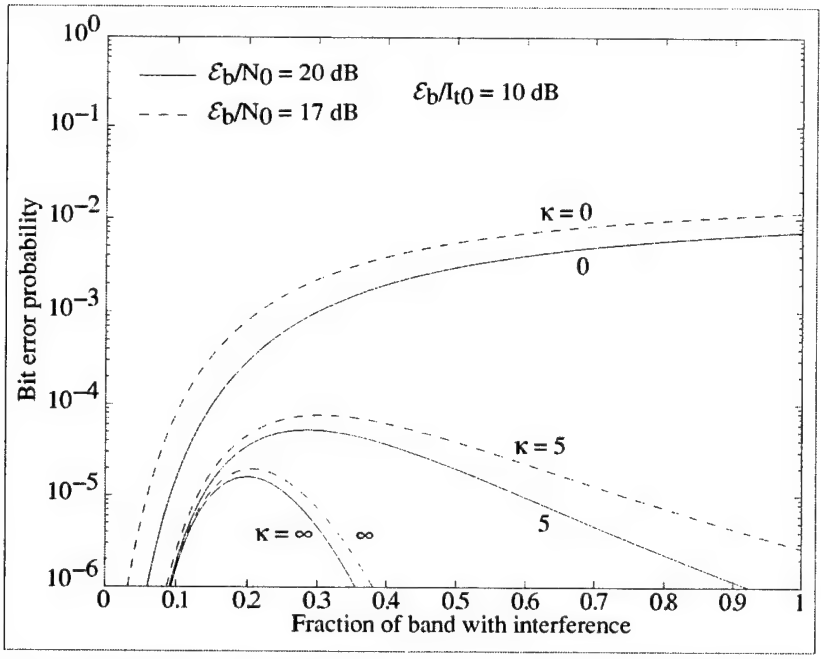


Figure 15. Performance of FH/MFSK with Reed-Solomon (32,12) code, nonbinary channel symbols, no erasures, and Ricean factor κ.

Much better performance against partial-band interference can be obtained by inserting erasures (Appendix A-1) among the demodulator output symbols before the symbol deinterleaving and hard-decision decoding. The decision to erase, which is made independently for each code symbol, is based on *side information*, which indicates which codeword symbols have a high probability of being incorrectly demodulated. The side information must be reliable so that only degraded symbols are erased, not correctly demodulated ones.

Side information may be obtained from known *test symbols* that are transmitted along with the data symbols in each dwell interval of a slow frequency-hopping signal [17]. A dwell interval during which the signal is in partial-band interference is said to be *hit*. If one or more of the N_t test symbols are incorrectly demodulated, then the receiver decides that a hit has occurred, and all codeword symbols in the same dwell interval are erased. Only one symbol of each codeword is erased if the interleaving ensures that only a single symbol of a codeword is in any particular dwell interval. Test symbols decrease the information rate, but this loss in negligible if $N_t << N_h$, which is assumed henceforth.

The probability of the erasure of a code symbol is

$$P_\epsilon = \mu P_{\epsilon 1} + (1-\mu)P_{\epsilon 0} \tag{6-10}$$

where $P_{\epsilon 1}$ is the erasure probability given that a hit occurred, and $P_{\epsilon 0}$ is the erasure probability given that no hit occurred. If a single incorrect demodulation of one of the N_t known test symbols causes an erasure, then

$$P_{\epsilon i} = 1 - (1 - P_{si})^{N_t}\ , \quad i = 0, 1 \tag{6-11}$$

where P_{s1} is the conditional channel-symbol error probability given that a hit occurred and P_{s0} is the conditional channel-symbol error probability given that no hit occurred.

A codeword symbol error can only occur if there is no erasure. Since test and codeword symbol errors are statistically independent when the partial-band interference is modeled as a white Gaussian process, the probability of a codeword symbol error is

$$P_s = \mu(1 - P_{\epsilon 1})P_{s1} + (1-\mu)(1 - P_{\epsilon 0})P_{s0} \tag{6-12}$$

and the conditional channel-symbol error probabilities are

$$P_{s1} = F\left(\frac{\mathcal{E}_s}{N_0 + I_{t0}/\mu}\right)\ , \quad P_{s0} = F\left(\frac{\mathcal{E}_s}{N_0}\right) \tag{6-13}$$

where (A-84) indicates that for MFSK symbols,

$$F(x) = \sum_{i=1}^{q-1} \frac{(-1)^{i+1}}{i+1}\binom{q-1}{i}\exp\left(-\frac{ix}{i+1}\right) \tag{6-14}$$

To account for Ricean fading, one must integrate (6-12) and (6-10) over the Ricean density (A-89). In the remainder of this section, we assume the absence of fading.

The word error probability for errors and erasures decoding is upper bounded in (A-35). Since most word errors result from decoding failures, it is reasonable to

assume that $P_b \lesssim P_w/2$. Therefore, the information-bit error probability is given by

$$P_b \approx \frac{1}{2}\sum_{j=0}^{n}\sum_{i=i_0}^{n-j}\binom{n}{j}\binom{n-j}{i}P_s^i P_\epsilon^j (1-P_s-P_\epsilon)^{n-i-j} \qquad (6\text{-}15)$$

where $i_0 = \max(0, \lceil (d_m - j)/2 \rceil)$ and $\lceil x \rceil$ denotes the smallest integer greater than or equal to x.

The P_b for FH/MFSK with $q = 5$, an extended Reed-Solomon (32,12) code, and errors and erasures decoding with $N_t = 2$ is shown in Figure 16. Fading is absent, and (6-10) to (6-15) are used. A comparison of Figure 16 with the $\kappa = \infty$ curves of Figure 15 indicates that when $\mathcal{E}_b/N_0 = 20$ dB, erasures provide nearly a 7 dB improvement in the required $\mathcal{E}_b/I_{t0}$ for $P_b = 10^{-5}$. The erasures also confer immunity to partial-band interference that is concentrated in a small fraction of the hopping band and decrease the sensitivity to $\mathcal{E}_b/N_0$.

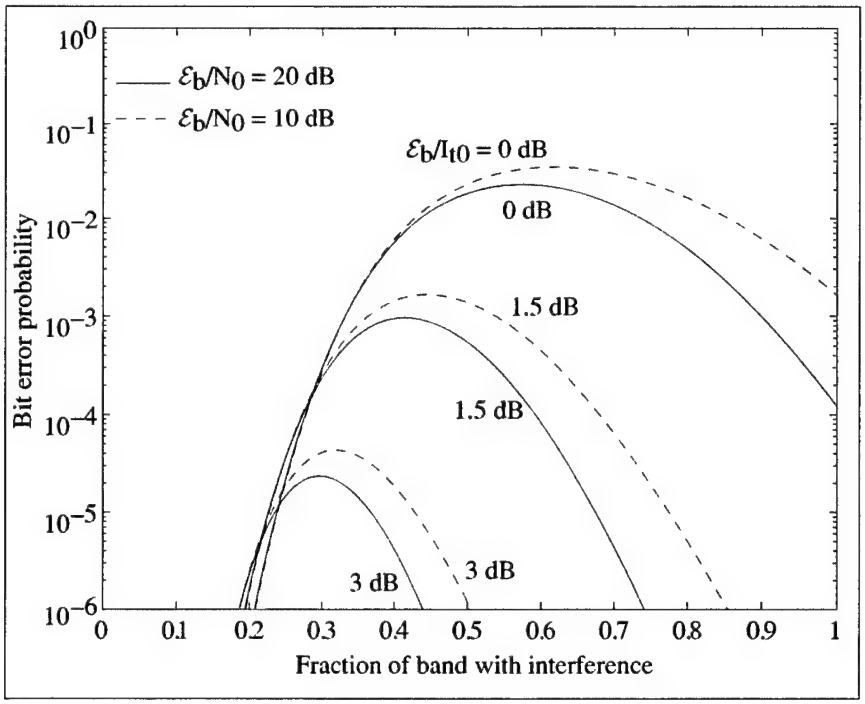


Figure 16. Performance of FH/MFSK with Reed-Solomon (32,12) code, nonbinary channel symbols, erasures, $N_t = 2$, and no fading.

There are other options for generating side information and, hence, erasure insertion in addition to demodulating test symbols. One might use a radiometer to measure the energy in the current frequency channel, a future channel, or an adjacent channel. Erasures are inserted if the energy is inordinately large. This method does not have the overhead cost in information rate that is associated with the use of test symbols. Other methods without overhead cost [17] use the soft information provided by the inner decoder of a concatenated code or use the

outputs of the parallel MFSK envelope detectors, which provide the decision variables applied to the MFSK decision device of Figure 12b.

The *output threshold test* (OTT) compares the largest decision variable to a threshold to determine whether the corresponding demodulated symbol should be erased. The *ratio threshold test* (RTT) computes the ratio of the largest decision variable to the second largest one. This ratio is then compared to a threshold to determine an erasure. If the values of both $\mathcal{E}_b/N_0$ and $\mathcal{E}_b/I_{t0}$ are known, then optimum thresholds for the OTT, the RTT, or a hybrid method can be calculated [18]. It is found that the OTT tends to outperform the RTT when $\mathcal{E}_b/I_{t0}$ is sufficiently low, but the opposite is true when $\mathcal{E}_b/I_{t0}$ is sufficiently high. If side information concerning the presence or absence of the partial-band interference is available at the receiver and if the interference power is high, then a threshold determined by $\mathcal{E}_b/N_0$ only and a separate threshold determined by $\mathcal{E}_b/(N_0 + I_{t0})$ can be used to further improve the performance of the errors and erasures decoding. The main disadvantage of the OTT and the RTT relative to the test-symbol method is the need to estimate $\mathcal{E}_b/N_0$ and either $\mathcal{E}_b/I_{t0}$ or $\mathcal{E}_b/(N_0 + I_{t0})$.

Proposed erasure methods are based on the use of MFSK symbols, and their performances against partial-band interference improve as the alphabet size q increases. For a fixed hopping band, the number of frequency channels decreases as q increases, thereby making an FH/MFSK system more vulnerable to narrowband jamming signals (Section 4.2) or multiple-access interference (Section 7.1). Thus, we examine alternatives that give less protection against partial-band interference in exchange for enhanced protection against multiple-access interference.

Figure 17 depicts P_b for FH/MFSK with $q = 3$, an extended Reed-Solomon (8,3) code, and $N_t = 4$. A comparison of Figures 17 and 16 indicates that reducing the alphabet size while preserving the code rate has increased the system sensitivity to $\mathcal{E}_b/N_0$, increased the susceptibility to interference concentrated in a small fraction of the hopping band, and raised the required $\mathcal{E}_b/I_{t0}$ for a specified P_b by 5 to 9 dB.

Another approach is to represent each nonbinary code symbol by a sequence of $m = \log_2 q$ consecutive binary channel symbols. Then an FH/MSK or FH/DPSK system can be implemented to provide a large number of frequency channels and, hence, better protection against multiple-access interference. Equations (6-10), (6-11), and (6-13) are still valid. However, since a code-symbol error occurs if any of its m component channel symbols is incorrect, (6-12) is replaced by

$$P_s = 1 - [1 - \mu(1 - P_{\epsilon 1})P_{s1} - (1 - \mu)(1 - P_{\epsilon 0})P_{s0}]^m \qquad (6\text{-}16)$$

and (6-14) is replaced by (6-4), where $\xi = 1/2$ for MSK and $\xi = 1$ for DPSK. The results for an FH/DPSK system with an extended Reed-Solomon (32,12) and $N_t = 10$ binary test symbols are shown in Figure 18. It is assumed that $N_h >> 1$ so that the loss due to the reference symbol in each dwell interval is negligible. The

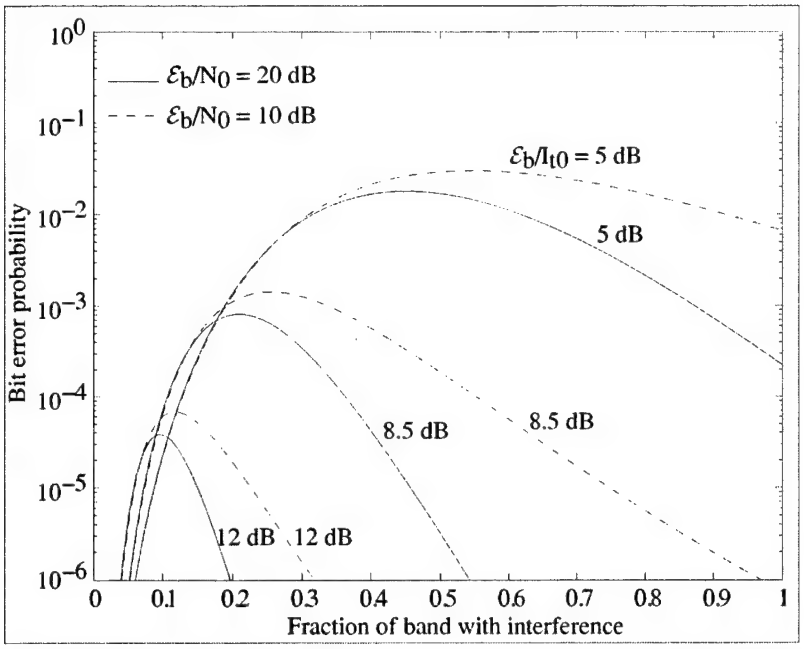


Figure 17. Performance of FH/MFSK with Reed-Solomon (8,3) code, nonbinary channel symbols, erasures, $N_t = 4$, and no fading.

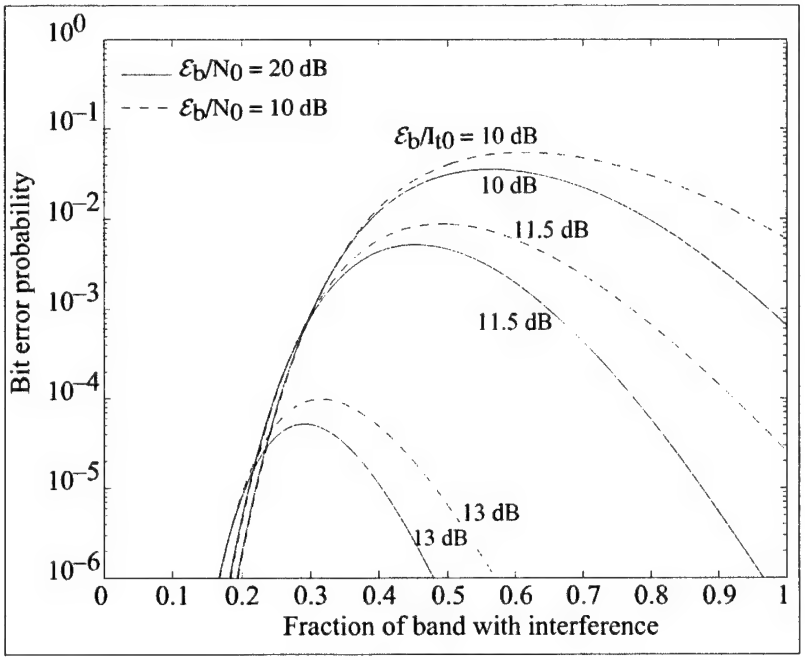


Figure 18. Performance of FH/DPSK with Reed-Solomon (32,12) code, binary channel symbols, erasures, $N_t = 10$, and no fading.

plots in Figure 18 are similar in form to those of Figure 16, but the transmission of binary rather than nonbinary symbols has caused approximately a 10 dB increase in the required $\mathcal{E}_b/I_{t0}$ for a specified P_b. Figure 18 is applicable to orthogonal FSK and MSK if $\mathcal{E}_b/I_{t0}$ and $\mathcal{E}_b/N_0$ are both increased by 3 dB to compensate for the lower value of ξ.

An alternative to erasures that uses binary channel symbols is an FH/DPSK system with concatenated coding, which has the form illustrated in Figures A-14 and A-15. Although generally unnecessary in a fast frequency-hopping system, the channel interleaver and deinterleaver may be required in a slow frequency-hopping system to ensure independent symbol errors at the decoder input. Consider a concatenated code comprising a Reed-Solomon (n, k) outer code and a binary convolutional inner code. The inner Viterbi decoder performs hard-decision decoding to limit the impact of individual symbol metrics. Assuming that $N_h >> 1$, the symbol error probability is given by (6-2) and (6-4) with $\xi = 1$. The probability of a Reed-Solomon symbol error, P_{s1}, at the output of the Viterbi decoder is upper bounded by (A-131) and (A-118). Setting $P_s = P_{s1}$ in (6-9) then provides an upper bound on P_b. Figure 19 depicts this bound for an outer Reed-Solomon (31,21) code and an inner rate-1/2, $K = 7$ convolutional code. This concatenated code provides a better performance than the Reed-Solomon (32,12) code with binary channel symbols, but a much worse performance than the latter code with nonbinary channel symbols.

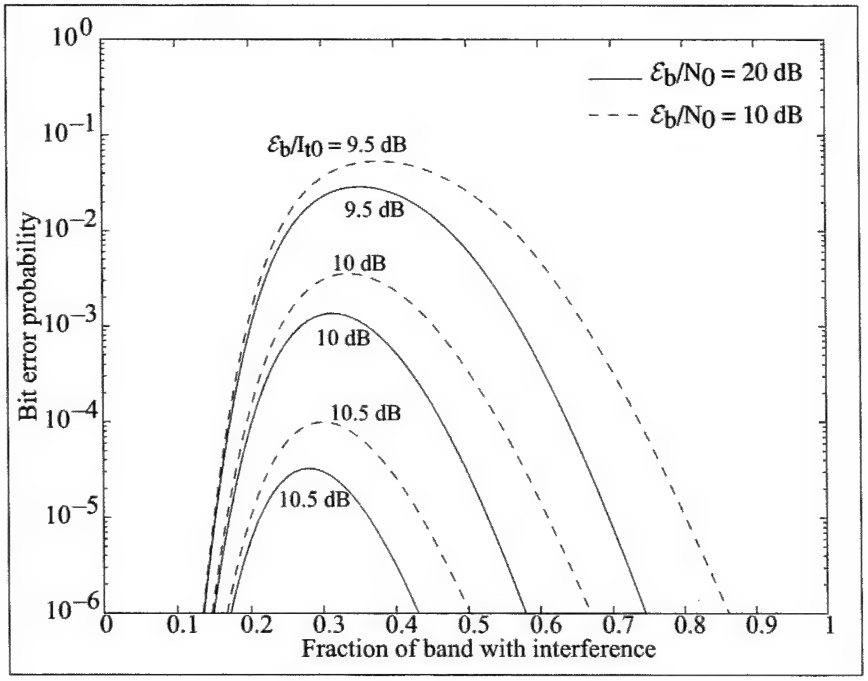


Figure 19. Performance of FH/DPSK with concatenated code, hard decisions, and no fading. Inner code is convolutional (rate = 1/2, K = 7) code and outer code is Reed-Solomon (31,21) code.

Figures 4 to 7 indicates that a reduction in the alphabet size for channel symbols increases the system susceptibility to partial-band interference. The primary reason is the reduced energy per channel symbol.

6.2 Trellis-Coded Modulation

Trellis-coded modulation is a combined coding and modulation method that is usually applied to coherent digital communications over bandlimited channels (Appendix A-2). Multilevel and multiphase modulations are used to enlarge the signal constellation while not expanding the bandwidth beyond what is required for the uncoded signals. Since the signal constellation is more compact, there is some modulation loss that detracts from the coding gain, but the overall gain can be substantial. Since a noncoherent demodulator is usually required for frequency-hopping communications, the usual coherent trellis-coded modulations are not suitable. Instead, the trellis coding may be implemented by expanding the signal set for $M/2$-ary MFSK to M-ary MFSK [19]. Although the frequency tones are uniformly spaced, they are allowed to be nonorthogonal to limit or avoid bandwidth expansion.

Trellis-coded 4-ary MFSK is illustrated in Figure 20 for a rate-1/2 code with four states. The signal set partitioning, shown in Figure 20(a), partitions the set of four signals or tones into two subsets, each with two tones. The partitioning doubles the frequency separation between tones from Δ Hz to 2Δ Hz. The mapping of code bits into signals is indicated. In Figure 20(b), the numerical labels denote the signal assignments associated with the state transitions in the trellis for a four-state encoder. The bandwidth of the frequency channel that accommodates the four tones is approximately $B = 4\Delta$.

There is a trade-off in the choice of Δ because a small Δ allows more frequency channels and thereby limits the effect of multiple-access interference or multitone jamming, whereas a large Δ tends to improve the system performance against partial-band interference. If a trellis code uses four orthogonal tones with spacing $\Delta = 1/T_b$, where T_b is the bit duration, then $B = 4/T_b$. The same bandwidth results when an FH/FSK system uses two orthogonal tones, a rate-1/2 code, and binary channel symbols since $B = 2/T_s = 4/T_b$. The same bandwidth also results when a rate-1/2 binary convolutional code is used and each pair of code symbols is mapped into a 4-ary channel symbol. The performance of the 4-state, trellis-coded, 4-ary MFSK frequency-hopping system [19] indicates that it is not as strong against worst-case partial-band interference as an FH/MFSK system with a rate-1/2 convolutional code and 4-ary channel symbols or an FH/FSK system with a Reed-Solomon (32,16) code and errors and erasures decoding. The advantage of trellis-coded modulation in a frequency-hopping system is its relatively low implementation complexity.

6.3 Turbo Codes

Turbo codes provide an alternative to errors and erasures decoding for suppressing partial-band interference. A turbo-coded frequency-hopping system that uses spectrally compact channel symbols will also resist multiple-access interference. An accurate estimate of the variance of the interference plus noise,

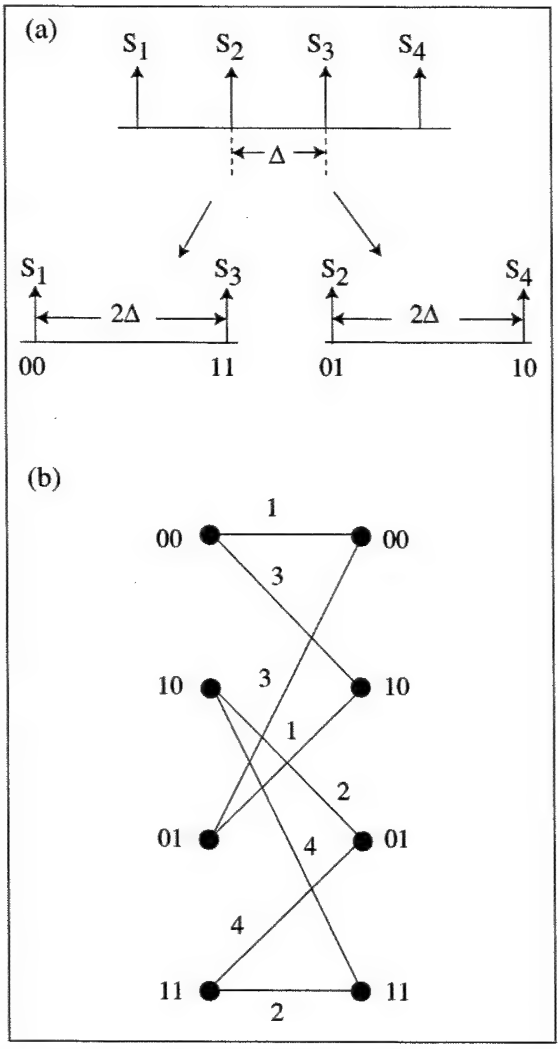


Figure 20. Rate-1/2, four-state trellis-coded 4-ary MFSK: (a) signal set partitioning and mapping of bits to signals, and (b) mapping of signals to state transitions.

which is modeled as zero-mean, white Gaussian noise, is always needed in the iterative turbo decoding algorithm (Appendix A-4). When the channel dynamics are much slower than the hopping rate, all the received symbols of a dwell interval may be used in estimating the variance associated with that dwell interval.

Consider an FH/DPSK system in which each code bit can take the values +1 or -1. The architecture of interactive turbo decoding and channel estimation is illustrated in Figure 21. The *log-likelihood ratio* (LLR) of a bit u_k conditioned on a received sequence $\mathbf{y} = (y_1\ y_2\ \ldots\ y_n)$ of demodulator outputs is defined as the natural logarithm of the ratio of the *a posteriori* probabilities:

$$L(u_k|\mathbf{y}) = \ln\left[\frac{P(u_k = +1|\mathbf{y})}{P(u_k = -1|\mathbf{y})}\right] \tag{6-17}$$

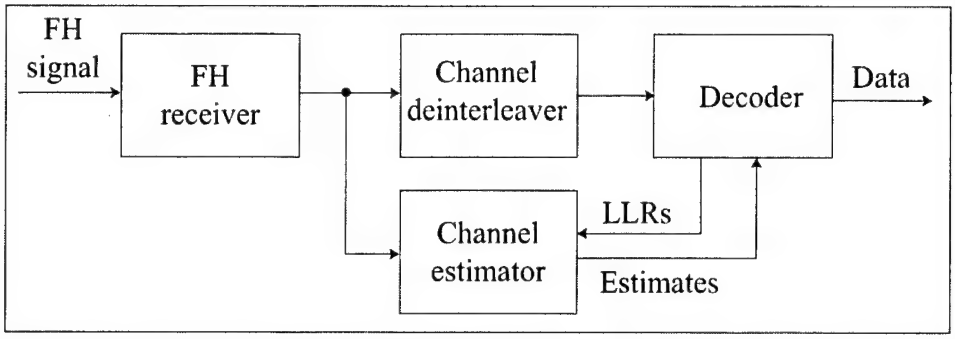


Figure 21. Receiver and decoder architecture for frequency-hopping system with turbo code.

Successive estimates of the LLRs of the code bits are computed by each component decoder during the iterative decoding of the turbo code. The usual turbo decoding is extended to include the iterative updating of the LLRs of both the information and parity bits. The fact that

$$P(u_k = -1|\mathbf{y}) = 1 - P(u_k = +1|\mathbf{y}) \tag{6-18}$$

and (6-17) imply that the *a posteriori* probabilities are

$$p_{1k} = P(u_k = +1|\mathbf{y}) = \frac{1}{1 + \exp[-L(u_k|\mathbf{y})]} \tag{6-19}$$

$$p_{0k} = P(u_k = -1|\mathbf{y}) = \frac{\exp[-L(u_k|\mathbf{y})]}{1 + \exp[-L(u_k|\mathbf{y})]} \tag{6-20}$$

Using these equations, the channel estimator can convert a LLR transferred after a component decoder iteration into estimates of p_{1k} and p_{0k}. The estimate of the variance of the interference plus noise during dwell interval j may be based on the N_h demodulator outputs during the dwell interval and is given by [20]

$$\hat{\sigma}_{tj}^2 = \frac{1}{N_h - 1} \sum_{k=0}^{N_h-1} [p_{0k}(y_k + 1)^2 + p_{1k}(y_k - 1)^2] + c \tag{6-21}$$

where y_k is the kth component of $\mathbf{y}$ during dwell interval j and c is a constant added to unbias the estimate. The set of estimates $\hat{\sigma}_{tj}^2$ for all the dwell intervals is sent to the other component decoder for the next decoding iteration.

A simulation of a turbo-coded FH/DPSK system [20] that uses (6-21) indicates that its performance is more than 2 dB better than that shown in Figure 16. The rate-1/3 turbo code uses two 4-state recursive systematic convolutional encoders, each with octal generator (5,7), a 200-bit turbo interleaver, ideal channel interleaving, 5 decoder iterations, $N_h = 10$, and $\mathcal{E}_b/N_0 = 20$ dB, which is known *a priori* by the receiver. After each iteration by a component decoder, its LLRs are updated and the extrinsic information is transferred to the other component decoder.

Estimates of the independent random carrier phase and the Rayleigh fading amplitude for each dwell interval can be integrated into the iterative decoding of a turbo code if these parameters are constants over the dwell interval [20], [21]. Bit estimates generated by one component decoder are used in computing parameter estimates that are passed to the other component decoder. For a sufficiently large dwell interval, the resulting performance is almost as good as theoretically possible with perfect side information about the carrier phase and the fading amplitude. Known symbols may be inserted into the transmitted code symbols to facilitate the estimation, but the energy per information bit is reduced. Increasing N_h improves the estimates because they may be based on more observations and more known symbols can be accommodated. However, since the reduction in the number of independent hops per information block of fixed size decreases the diversity, and hence the independence of errors, there is a limit on N_h beyond which a performance degradation occurs.

Although turbo codes are generally used with binary channel symbols, their error-correcting capability is strong enough to compensate for the relatively low channel-symbol energy. However, if the system latency and computational complexity of turbo codes is unacceptable, then there is a tradeoff in the choice of the modulation and code.

7. Frequency-Hopping Multiple-Access Networks

Frequency-hopping systems are usually part of a *frequency-hopping code-division multiple-access* (FH/CDMA) network in which all systems share the same M frequency channels. In a *synchronous FH/CDMA* network, the systems coordinate their frequency transitions and hopping patterns. Consequently, as many as M frequency-hopping signals can be simultaneously accommodated by the network with insignificant multiple-access interference at any of the active receivers. Network coordination is much simpler to implement than for a direct-sequence CDMA network because the timing alignments must be within a small fraction of a hop duration instead of a small fraction of a spreading-sequence chip. Multipath signals and errors in range estimates can be accommodated at some cost in the energy per information bit by increasing the switching time between frequency-hopping pulses. However, some type of centralized or cellular architecture is required, and such an architecture is often unavailable.

7.1 Asynchronous FH/CDMA Networks

An *asynchronous FH/CDMA* network has systems that transmit and receive autonomously and asynchronously. When two or more frequency-hopping signals using the same frequency channel are received simultaneously, they are said to *collide*. Since the probability of a collision in an asynchronous network is decreased by increasing the number of frequency channels in the hopset, it is highly desirable to choose a data modulation that has a compact spectrum. Good candidates are FH/CPFSK systems that use a frequency discriminator for demodulation. Binary CPFSK with $h = 0.7$ and $BT_s = 1$ provides excellent potential performance if the spectral splatter and intersymbol interference generated by this modulation are negligible. However, for approximately the same degree of spectral splatter and intersymbol interference as MSK with $BT_s = 1$, the bandwidth must be increased so that $BT_s = 1.4$, which reduces the number of frequency channels M in a fixed hopping band. This much reduction in M is enough to completely offset the intrinsic performance advantage of binary CPFSK with $h = 0.7$. Thus, the choice between the latter and MSK will depend on the details of the impact of the spectral splatter and intersymbol interference.

Let d represent the *duty factor*, which is defined as the probability that an interferer using the same frequency channel will degrade the reception of a symbol. Thus, $d = q_1 q_2$ is the product of the probability q_1 that an interferer is transmitting and the probability q_2 that a significant portion of the interferer's transmitted waveform occurs during the symbol interval. The probability q_2 is

upper bounded and well approximated by the probability that there is any overlap in time of the interference and the symbol interval. For synchronous frequency hopping, $q_2 = 1$. Since $T_{sw} > T_s$, it follows from elementary probability that for asynchronous frequency hopping, $q_2 \approx (T_d + T_s)/T_h$. For voice communications with voice-activity detection, $q_1 = 0.4$ is a typical value. Assuming that an interferer may transmit in any frequency channel with equal probability, the probability that a potentially interfering signal collides with the desired signal during a symbol interval is

$$c = \frac{d}{M} \tag{7-1}$$

When a collision occurs, the symbol is said to be *hit* by the interfering signal. For MFSK, M is given by (6-7).

Consider an FH/CDMA network of $N + 1$ asynchronous systems with negligible spectral splatter and intersymbol interference. The code symbols are interleaved so that each code symbol of a codeword is transmitted in a separate dwell interval. The switching time $T_{sw} > T_s$ so that each potentially interfering frequency-hopping signal uses at most one frequency channel during the reception of one symbol of a desired signal. The $2N_t$ test symbols are split into separate sets of $N_t \geq 1$ test symbols at each end of a dwell interval [17]. Thus, if a code symbol is hit by one or more of the N interfering signals, then at least one set of the test symbols in that same dwell interval is also hit. For analytical simplicity, we make the following assumptions:

1. If at least one of the two test symbols at the opposite ends of a dwell interval is hit, then an erasure is always made. Thus, if a code symbol is hit, an erasure is always made.
2. If a code symbol is not hit, then this condition has a negligible influence on the probability that one of the two end test symbols is hit.
3. The probability that both end test symbols are hit is negligible.

These assumptions are approximately valid if $N_h >> N_t \geq 1$ and the N interfering signals have approximately the same or more power than the desired signal.

The first assumption implies that the probability of the erasure of a code symbol is

$$P_\epsilon = [1 - (1 - c)^N] + (1 - c)^N P_{\epsilon 0} \tag{7-2}$$

where $P_{\epsilon 0}$ is the erasure probability given that no hit of the code symbol occurred. Observe that if neither of the end test symbols is hit, then no test symbol is hit. Therefore, the assumptions imply that

$$P_{\epsilon 0} = 2[1 - (1 - c)^N] + [2(1 - c)^N - 1]\left\{1 - \left[1 - F\left(\frac{\mathcal{E}_s}{N_0}\right)\right]^{2N_t}\right\} \tag{7-3}$$

where the first term is the probability that one of the two end test symbols is hit, and the term in braces is the probability that although no test symbols are hit, an erasure occurs because at least one of the detected test symbols is incorrect. For MFSK modulation and no fading, each channel symbol is a code symbol and $F(x)$ is given by (6-14). Under the first assumption, the code-symbol error probability is

$$P_s = (1-c)^N(1-P_{\epsilon 0})F\left(\frac{\mathcal{E}_s}{N_0}\right) \tag{7-4}$$

Suppose that each code symbol is represented by $m = \log_2 q$ binary channel symbols that are interleaved over m dwell intervals. Since all m channel symbols must be received correctly for there to be no code-symbol error, then the independence of channel-symbol errors implies that

$$P_s = 1 - \left[1-(1-c)^N(1-P_{\epsilon 0})F\left(\frac{\mathcal{E}_s}{N_0}\right)\right]^m \tag{7-5}$$

where $F(x)$ is given by (6-4) with ξ = 1/2 for orthogonal FSK or MSK and ξ = 1 for DPSK. Equation (6-15) gives P_b for errors and erasures decoding. The fundamental advantage of MSK is the reduced bandwidth per frequency channel. Thus, instead of (6-7),

$$M = \left\lfloor \frac{rW}{B_{MSK}} \right\rfloor \geq \left\lfloor \frac{2rW}{B_u} \right\rfloor \tag{7-6}$$

since $B_{MSK} \leq B_u/2$.

Figure 22 illustrates P_b versus N for FH/MFSK, FH/FSK, FH/DPSK, and FH/MSK systems that use a Reed-Solomon (32,12) code with errors and erasures decoding against asynchronous multiple-access interference. The plots are computed from (7-2) to (7-5) with M given by (6-7) for MFSK and orthogonal FSK, and the lower bound in (7-6) for MSK. In all cases, W/B_u = 1000 and d = 1. The 5-ary MFSK channel symbols have N_t = 2 while the systems that use binary channel symbols have N_t = 10. The huge benefit obtained from using binary channel symbols and the further substantial benefit from using MSK are apparent in the figure. These benefits must be weighed against the disadvantage of binary channel symbols in the presence of partial-band interference, as shown in Section 6. The figure illustrates that as $\mathcal{E}_b/N_0$ drops from 17 dB to 14 dB, the FH/MSK systems degrade substantially while the FH/MFSK systems degrades imperceptibly.

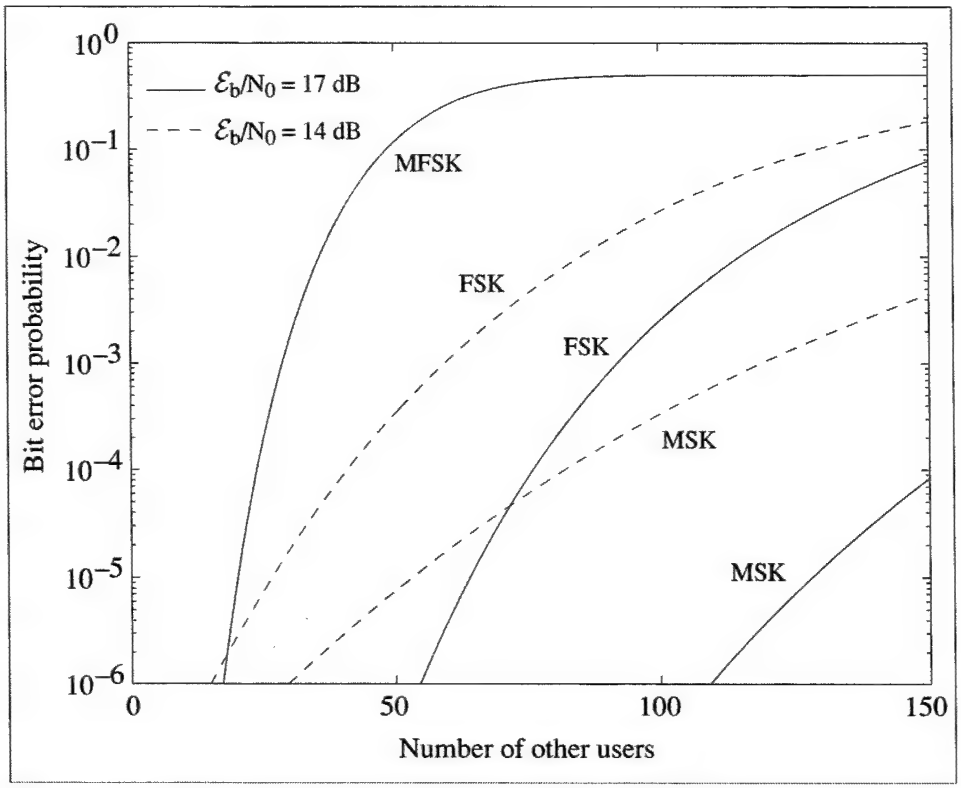


Figure 22. Performance of frequency-hopping systems with Reed-Solomon (32,12) code, various modulation, erasures, $W/B_u = 1000$, $d = 1$, and no fading. $N_t = 10$ for binary modulations; $N_t = 2$ for 5-ary MFSK.

To obtain good performance against both partial-band interference and multiple-access interference, a turbo code and binary channel symbols are needed. However, even if $\mathcal{E}_s$ is known, perhaps through power control, the turbo decoder computation must be modified to account for the fluctuations from symbol-to-symbol in the interference-plus-noise variance caused by multiple-access interference [20]. When DPSK is the modulation, suitable modifications use (A-141) and (A-134).

If a turbo code is not feasible, then a Reed-Solomon code with errors and erasures decoding is a good choice. However, a tradeoff is necessary in the choice of the modulation. If one is primarily interested in avoiding multiple-access interference, then binary channel symbols are desirable. If stronger protection against partial-band interference but weaker protection against multiple-access interference is needed, then nonbinary channel symbols are preferable.

7.2 Mobile Peer-to-Peer and Cellular Networks

The two principal types of CDMA are direct-sequence CDMA and frequency-hopping CDMA. Two major advantages of frequency hopping are that it can be implemented over a much larger frequency band than it is possible to implement direct-sequence spreading, and that the band can be divided into noncontiguous

segments. Another major advantage is that frequency hopping provides resistance to multiple-access interference, while not requiring power control to prevent the near-far problem. In direct-sequence systems, accurate power control is crucial but becomes much less effective as the fading rate increases. These advantages of frequency hopping will be decisive in many applications.

Mobile peer-to-peer communications are used in mobile communication networks that possess no supporting infrastructure, fixed or mobile; each user has identical signal processing capability. Peer-to-peer communications have both commercial applications and important military applications, the latter primarily because of their robustness in the presence of node losses. Mobile frequency-hopping CDMA systems [22] are suitable for both peer-to-peer and cellular communication networks. Power control and, hence, current direct-sequence CDMA are not viable for peer-to-peer communications because of the lack of a centralized architecture. Current plans to use multiuser detection in direct-sequence CDMA systems still require power control, which is highly desirable for the synchronization.

Compared with the alternatives, CDMA is advantageous for cellular networks because it eliminates the need for frequency and timeslot coordination among cells, allows complete frequency reuse in all cells, and can fully exploit intermittent voice signals and sectorization. Both frequency-hopping and direct-sequence systems are viable choices for mobile cellular CDMA communications.

Through analysis and simulation, a unified evaluation of the potential performance of both mobile peer-to-peer and sectorized cellular frequency-hopping CDMA systems is presented. The equivalent number of frequency channels and the minimum signal-to-noise ratio (SNR) are defined and shown to be important parameters in understanding and predicting network capacity. The effects of spectral splatter are analyzed. Separated orthogonality is defined and shown to be useful in completely eliminating intracell or intrasector interference. Spatial diversity by postdetection rather than predetection combining is proposed and shown to be invaluable. Noncoherent demodulation by a frequency discriminator rather than parallel matched filters and envelope detectors is proposed and shown to be effective. It is shown that even without exploiting either its natural bandwidth advantage or power control, frequency-hopping CDMA provides an uplink capacity nearly the same as direct-sequence CDMA with realistic power-control imperfections.

The propagation path losses are modeled as the result of power-law losses, shadowing, and fading. In the absence of shadowing and fading, the received signal power at an omnidirectional antenna for communications over a fixed range r has an average value called the *area-mean power*, where the average is calculated over a specified geographic area. The power-law model assumes that the area-mean power has the form

$$p_a = p_0 \left(\frac{r}{R_0} \right)^{-\beta} \tag{7-7}$$

where p_0 is the power when the range is $r = R_0$, and β is the power law. A typical value of the power law for an urban area is $\beta = 4$. In the absence of fading, the received *local-mean power* fluctuates about the area-mean power due to *shadowing*, which is the effect of path-to-path differences in terrain profiles for paths of the same distance. On the basis of extensive empirical data, the local-mean power p_l is assumed to be lognormally distributed, which implies that

$$p_l = p_a 10^{\xi/10} \tag{7-8}$$

where ξ is a zero-mean random variable with a normal distribution. The standard deviation of ξ is denoted by σ_s, which is expressed in decibels and is assumed to be the same for all signals. A straightforward derivation using (7-7) and (7-8) indicates that the probability distribution function of the normalized local-mean power p_l/p_0 is

$$F(x) = 1 - \frac{1}{2} \operatorname{erfc} \left\{ \frac{a}{\sigma_s} \ln \left[x \left(\frac{r}{R_0} \right)^{\beta} \right] \right\} \tag{7-9}$$

where erfc{ } denotes the complementary error function and $a = (10 \log_{10} e)/\sqrt{2}$. The fading causes a power fluctuation about the local-mean power.

It is assumed that L omnidirectional antennas are deployed to achieve spatial diversity at the mobiles. The antennas are separated from each other by several wavelengths, so that the fading of both the desired signal and the interfering signals at one antenna is independent of the fading at the other antennas. A few wavelengths are adequate because mobiles, in contrast to base stations, tend to receive superpositions of reflected waves arriving from many random angles. Because of practical physical constraints, spatial diversity will ordinarily be effective only if the carrier frequencies exceed roughly 1 GHz. Polarization diversity and other forms of adaptive array processing are alternatives.

One method of combining antenna outputs is predetection combining, which requires the estimation of the signal and interference-plus-noise power levels at each antenna for maximal-ratio combining or selection diversity and requires the cophasing of the L antenna outputs for maximal-ratio or coherent equal-gain combining. Since the relative phases and power levels of the signals at the L antennas change after every hop, it is almost always impractical to implement predetection combining. As a much more practical alternative, a receiver can combine the demodulated outputs rather than the signals from the L antennas. This postdetection combining eliminates the cophasing and does not require the time alignment of L signals in practical applications because any misalignment is much smaller than a symbol duration. The estimation of power levels can be

eliminated by the use of a fixed combining rule, such as equal-gain or square-law combining.

In the receiver of a frequency-hopping system, each antenna output is dehopped and filtered. The interference plus noise in each dehopped signal is approximated by independent bandlimited white Gaussian noise, with equivalent power given by

$$\sigma_1^2 = \sigma_n^2 + \sum_{i=1}^{K} p_{ui} \tag{7-10}$$

where σ_n^2 is the thermal noise power, K is the number of active interfering mobiles, and p_{ui} is the local-mean interference power received from mobile i. The Gaussian model is reasonable, especially for large numbers of mobiles, because the interference signals are asynchronous, fade independently, and experience different Doppler shifts. The total interference power is approximately uniform (white) over the receiver passband following dehopping if $BT_s = \zeta \leq 1$. The L diversity antennas are assumed to be close enough to each other that the power-law losses and shadowing are nearly the same, and thus the local-mean power from a mobile is the same at each antenna. Each active interfering mobile may actually represent a cluster of mobiles. In this cluster, some discipline such as carrier-sense multiple access is used to ensure that there is at most one transmitted signal at any time.

The desired signal is assumed to experience frequency-nonselective Rayleigh fading. The Rayleigh fading model is appropriate under the pessimistic assumption that the propagation paths are often obstructed, and thus, the power of the direct line-of-sight signal is small compared with the reflected signal power. Frequency-nonselective fading occurs if $B < B_{\text{coh}}$. Rayleigh fading may be negligible if mobile speeds are very low, which would occur if each mobile consisted of a person walking. Shadowing would still occur but would be slowly varying over time.

Spectrally compact CPFSK signals do not have enough frequency shift to be demodulated by classical noncoherent demodulators with parallel matched filters and envelope detectors, but can be demodulated by a frequency discriminator. We consider binary MSK with discriminator demodulation. For postdetection diversity, the outputs of L discriminators are weighted and combined. The weighting is by the square of the envelope at the input to each discriminator. When the desired signal undergoes independent Rayleigh fading at each antenna and the channel parameters remain constant for at least one symbol duration, a calculation using the results of Adachi and Parsons [23] yields the symbol-error probability

$$P_s = \binom{2L-1}{L} \left(\frac{1}{4} + \frac{1}{3}\zeta^2\right)^L (\bar{\rho})^{-L} \tag{7-11}$$

where $\zeta = BT_s$, $\bar{\rho} = p_s/\sigma_1^2$, and p_s is the local-mean power of the desired signal. The information-bit error rate following hard-decision decoding can be calculated from P_s with well-known equations. The theoretical loss due to using postdetection rather than predetection combining is less than a decibel [23].

7.3 Peer-to-Peer Simulation Results

This section considers a peer-to-peer network of independent, identical, frequency-hopping systems that have omnidirectional antennas, generate the same output power, share the same carriers and frequency channels, and are nearly stationary in location over a single symbol duration. Since for peer-to-peer communications it is assumed that an interfering mobile may transmit in any frequency channel with equal probability, the probability that power from an interferer enters the transmission channel of the desired signal is

$$P_t = \frac{d}{M} . \tag{7-12}$$

It is assumed that M is sufficiently large that we may neglect the fact that a channel at one of the ends of the hopping band has only one adjacent channel instead of two. Consequently, the probability that the power from an interferer enters one of the two adjacent channels of the desired signal is

$$P_a = \frac{2d}{M} . \tag{7-13}$$

The probability that the power enters neither the transmission channel nor the adjacent channels is $(1 - 3d/M)$. These equations make it apparent that the performance of a frequency-hopping system depends primarily on the ratio $M_1 = M/d$. This ratio is called the *equivalent number of channels* because any decrease in the duty factor has the same impact as an increase in the number of frequency channels; what matters most for performance is this ratio.

In the simulation, the locations of the mobiles are assumed to be uniformly distributed in a circular region surrounding a specific mobile receiver, as illustrated in Figure 23. Therefore, the radial distance of a mobile from the receiver has the probability distribution function

$$G(r) = \frac{r^2}{R^2} , \quad 0 < r \leq R \tag{7-14}$$

where R is the radius of the circle. The distance of the desired mobile is randomly selected according to this distribution with $R = R_0$, where R_0 is the maximum communication range and corresponds to a received area-mean signal power equal to p_0. The distance of each interfering mobile is randomly selected according to this distribution with $R = R_1$. The selected distance of the desired mobile is substituted into (7-9) as the value of r, and then (7-9) is used to randomly select the- local-mean power of the desired signal at the receiver. The

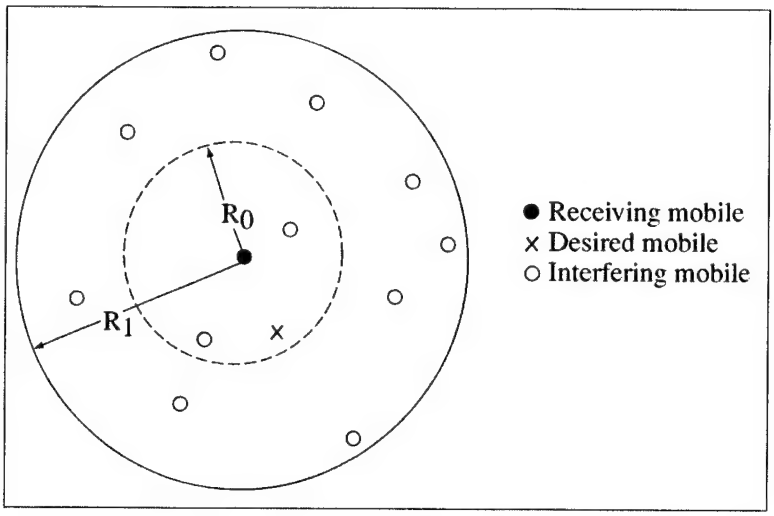


Figure 23. Geometry of a peer-to-peer communication network.

probabilities given by (7-12) and (7-13) are used to determine if an interfering mobile produces power in the transmission channel or in one of the adjacent channels of the desired signal. If the power enters the transmission channel, then the power level is randomly selected according to (7-9) with the distance of the mobile substituted. If the power enters one of the adjacent channels, then the potential local-mean power level is first randomly selected via (7-9) and then multiplied by K_s to determine the net interference power p_{ui} that appears in (7-10). The effects of p_0 and σ_n^2 are determined solely by the *minimum area-mean SNR*, which occurs at the maximum range $r = R_0$ of the desired signal and is equal to p_0/σ_n^2. Once the local-mean power levels and the noise power are calculated, the symbol error probability P_s is calculated with (7-10) and (7-11). Each simulation experiment was repeated for 10,000 trials, with different randomly selected mobile locations in each trial. The performance measure is the *spatial reliability*, which is defined as the fraction of trials for which P_s is less than a specified performance threshold E. The appropriate value of the threshold depends on the desired information-bit error probability and the error-correcting code. The spatial reliability is essentially the probability that an outage does not occur.

Figures 24 to 26 depict the results of three simulation experiments for peer-to-peer networks. The figures plot the spatial reliability as a function of K for various values of L, assuming (7-11), Rayleigh fading, MSK, and that $\beta = 4$, $\sigma_s =$ 8 dB, $E = 0.01$, $\zeta = 1$, $K_s = 0.015$, $R_0 = 1$, and $R_1 = 2$. The value of K_s results from assuming contiguous frequency channels with center frequencies separated by B. The units of R_0 and R_1 are immaterial to the calculation of the spatial diversity.

Figure 24 provides a baseline with which the other figures may be compared. For this figure, the assumptions are that $M_1 = 250$, and the minimum area-mean SNR = 20 dB. The number of equivalent frequency channels M_1 could model voice

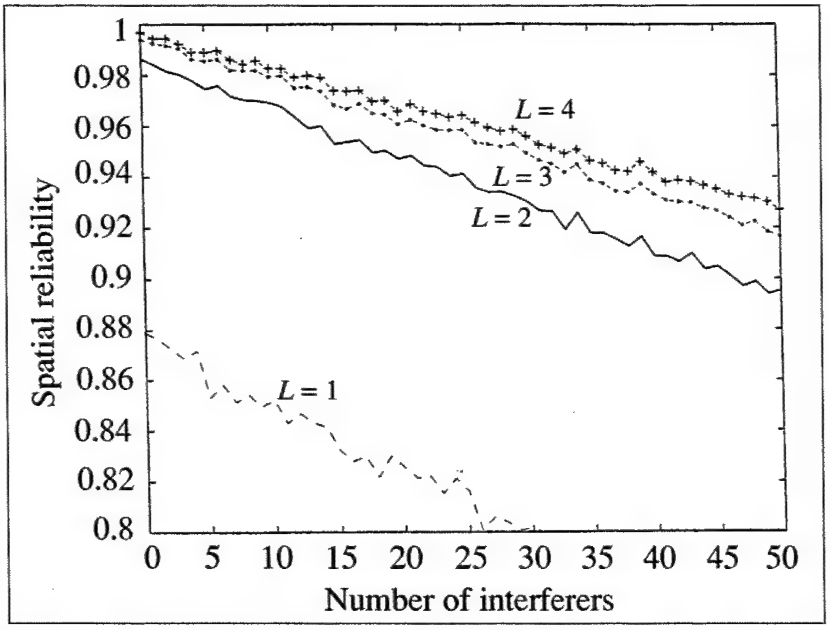


Figure 24. Spatial reliability for $M_1 = 250$ and minimum area-mean SNR = 20 dB.

communications with $M = 90$ channels and $d = 0.36$; alternatively, it could model continuous data communications with $M = 225$ and $d = 0.9$. The figure illustrates the dramatic performance improvement provided by dual spatial diversity when Rayleigh fading occurs. Further increases in diversity yield diminishing returns. One can assess the impact of the spectral splatter in this example by setting $K_s = 0$ and observing the change in the spatial reliability. The change is small, and nearly imperceptible if $K < 25$.

Figure 25 illustrates the effect of increasing the number of equivalent channels to $M_1 = 500$. Let the *capacity* of the network be defined as the maximum number of interfering mobiles for which the spatial reliability exceeds 0.95. Figures 24 and 25 and other simulation results indicate that for the parameter values selected, the capacity C for dual spatial diversity is approximately proportional to M_1; specifically, $C \approx 0.07\ M_1$ for $100 \leq M_1 \leq 1000$. If E is increased to 0.02, the capacity for dual spatial diversity increases by approximately 20 percent.

Figure 26 illustrates the sensitivity of the network to an increase in the minimum area-mean SNR, which may be due to a change in p_0 or σ_n^2. For no spatial diversity or dual diversity, a substantial performance improvement occurs when the minimum area-mean SNR = 25 dB. Other simulation results indicate that a decrease in the minimum area-mean SNR below 20 dB severely degrades performance.

Since (7-11) relates P_s to $\bar{\rho}$, the local-mean signal-to-interference-plus-noise ratio (SINR), the spatial reliability has an alternative and equivalent definition as the fraction of trials for which the SINR exceeds a specified threshold Z_l. Thus, the curves labeled $L = 1, 2, 3$, and 4 in Figures 24 to 26 (and later in Figures 28 to 31 correspond to $Z_l = 17.7$ dB, 10.0 dB, 7.7 dB, and 6.5 dB, respectively.

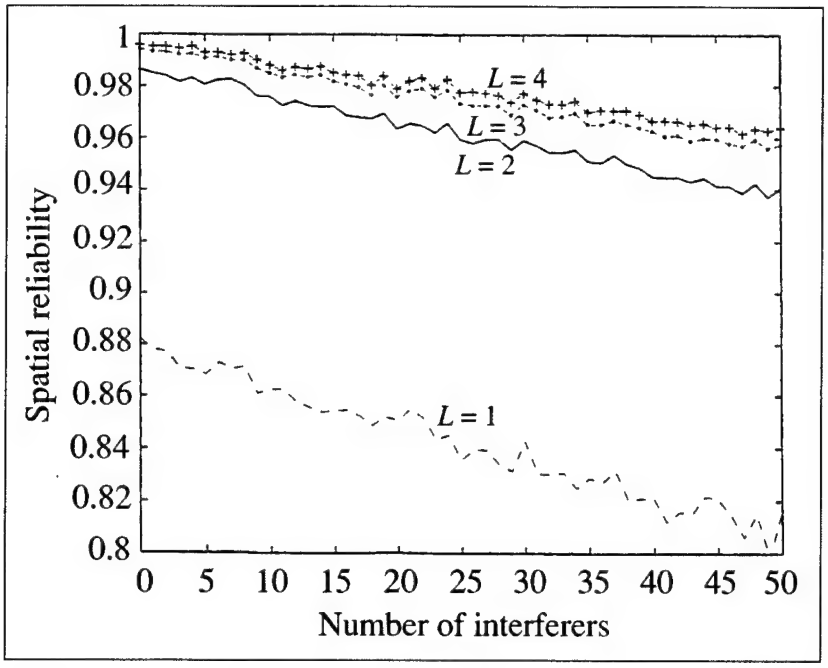


Figure 25. Spatial reliability for $M_1 = 500$ and minimum area-mean SNR = 20 dB.

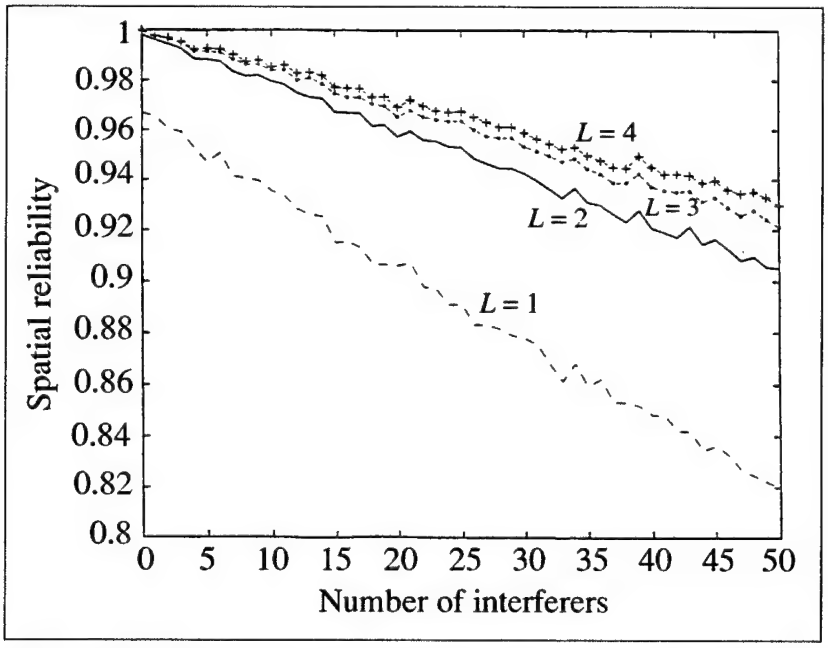


Figure 26. Spatial reliability for $M_1 = 250$ and minimum area-mean SNR = 25 dB.

7.4 Cellular Systems

In a cellular network, each base station assigns separate directional sector antennas or separate outputs of a phased array to cover disjoint angular sectors in both the transmitting and receiving modes. Typically, there are three sectors, and $2\pi/3$ radians are in each angular sector. The mobile antennas are assumed to be omnidirectional. Ideal sector antennas have uniform gain over the covered sector

and negligible sidelobes. With these antennas, only mobiles in the covered sector can cause multiple-access interference on an uplink from a mobile to a base station, and the number of interfering signals on the link is reduced by a factor s approximately equal to the number of sectors. Only the antenna serving a cell sector oriented toward a mobile can cause multiple-access interference on a downlink from the controlling base station to a mobile. Therefore, the number of interfering signals is reduced approximately by a factor s on both the uplinks and downlinks. Practical sector antennas have patterns with sidelobes that extend into adjacent sectors, but the performance degradation due to overlapping sectors is significant only for a small percentage of mobile locations. Ideal sector antennas are assumed in the subsequent simulation.

Spatial diversity may be obtained through the deployment of L antennas in each mobile and L antenna elements for each sector antenna of each base station. The antennas are separated from each other enough that the fading of both the desired signal and the interfering signals at one antenna is independent of the fading at the other antennas. A few wavelengths are adequate for a mobile because it tends to receive superpositions of reflected waves arriving from many random angles. Many wavelengths separation may be necessary for a base station located at a high position, and polarization diversity may sometimes be a more practical means of obtaining diversity.

In a cellular network, the frequency-hopping patterns can be chosen so that at any given instant in time, the frequencies of the mobiles within a cell sector are all different and, hence, the received signals are all orthogonal if the mobile transmissions are properly synchronized. Exact synchronization on a downlink is possible because a common timing is available. The advancing or retarding of the transmit times of the mobiles enables the arrival times at the base station of the uplink signals to be synchronized. The switching time or guard time between frequency-hopping pulses must be large enough to ensure that neither a small synchronization error nor multipath signals can subvert the orthogonality. The appropriate transmit times of a mobile can be determined from position information provided by the Global Positioning System and the known location of the base station. Alternatively, the transmit times can be determined from arrival-time measurements at the base station that are sent to the mobile. These measurements may be based on the adaptive thresholding [24] of the leading and/or trailing edges of a sequence of frequency-hopping pulses.

Let N_s denote the number of mobiles assigned to a cell sector. To ensure orthogonality of N_s received signals within a cell sector, a simple procedure is to generate a periodic frequency-hopping pattern that does not repeat until all the carrier frequencies in a hopset of size $M \geq N_s$ have been used. Mobile n is assigned this pattern with a delay of $n-1$ hop durations, where $n = 1, 2, \ldots, N_s$. If the patterns associated with different sectors are all drawn from a set of one-coincidence sequences [25], then any two signals from different cells or sectors will collide in frequency at a base station at most once during the period

of the hopping patterns. However, the use of one-coincidence sequences throughout a network requires frequency planning, which may be too costly in some applications.

It is possible to ensure not only the orthogonality of N_s signals in a sector but also that the received carrier frequencies in any two patterns are separated by at least νB, where ν is a positive integer, so that the spectral splatter is greatly reduced or negligible. Let $k = 0, 1, 2, \ldots, M-1$ label the hopset frequencies in ascending order. Suppose that a frequency-hopping pattern is generated that does not repeat until all the carrier frequencies in a hopset of size $M \geq \nu N_s$ have been used. When mobile 1 hops to frequency k, mobile n hops to frequency $[k + \nu (n-1)]$ modulo M. Frequency-hopping signals that use frequencies determined by this procedure are called *separated orthogonal signals*. Choosing $\nu = 2$ will generally be adequate because spectral splatter from channels that are not adjacent will be nearly always insignificant if a spectrally compact data modulation is used.

Frequency-hopping CDMA networks largely avoid the near-far problem by continually changing the carrier frequencies so that frequency collisions become brief, unusual events. Thus, power control in a frequency-hopping CDMA network is unnecessary, and all mobiles may transmit at the same power level. When power control is used, it tends to benefit signals from mobiles far from an associated sector antenna, while degrading signals from mobiles close to it. Simulation results [26] indicate that even perfect power control typically increases system capacity by only a small amount. There are good reasons to forego this slight potential advantage and not use power control. The required overhead may be excessive. If geolocation of mobiles is done by using measurements at two or more base stations, then the power control may result in significantly less signal power arriving at one or more base stations and the consequent loss of geolocation accuracy.

Consider communications between a base station and a mobile assigned to sector A of a particular cell, as illustrated in Figure 27 for a hexagonal grid of cells. Because of orthogonality, no other signal in sector A will use the same carrier frequency at the same time and thereby cause interference in the transmission channel (current frequency channel) of either the uplink or downlink. Consider another sector covered by the sector antenna of sector A; an example is sector B. Assuming that an interfering signal may independently use any frequency in the network hopset with equal probability, the probability that a mobile in the covered sector produces interference in the transmission channel of the uplink and degrades a particular symbol is

$$P_m = \frac{dN_s}{M} . \tag{7-15}$$

This equation also gives the probability that a sector antenna serving another sector that is oriented toward the desired mobile degrades a symbol by producing interference in the transmission channel of the downlink. Because of

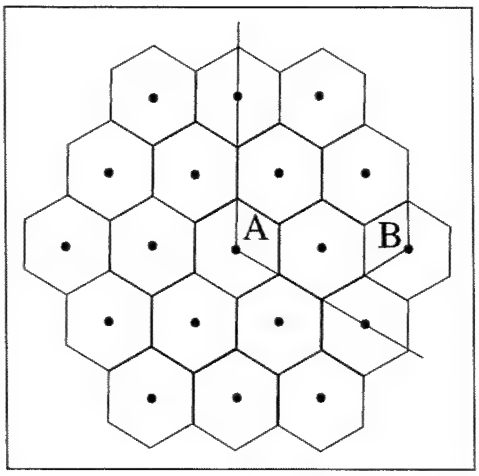


Figure 27. Hexagonal grid of cells. Communicators are in sector A. Sector B is an interfering sector.

orthogonality within each sector, no more than one signal from a sector will produce interference in the transmission channel of either link. A sector with mobiles that may interfere with communications over an uplink or a sector with an antenna that may produce interference over a downlink is called an *interfering sector*.

It is assumed that M is sufficiently large that we may neglect the fact that a channel at one of the ends of the hopping band has only one adjacent channel within the band instead of two. Let $N_1 = 1$ if a signal from an interfering sector uses the transmission channel of communicators in sector A; let $N_1 = 0$ if it does not. The probability that $N_1 = 1$ is N_s/M. The $N_s - N_1$ interference signals from a sector that do not enter the transmission channel are assumed to be randomly distributed among the $M - 1$ frequency channels excluding the transmission channel. There are $\binom{M-1}{N_s-N_1}$ ways to choose the channels with interference signals. There are $\binom{2}{1}$ ways to choose one of the two adjacent channels to have an interference signal and $\binom{M-3}{N_s-N_1-1}$ ways to choose $N_s - N_1 - 1$ channels with interference signals out of the $M - 3$ channels excluding both the transmission channel and the adjacent channels. The probability that an adjacent channel with an interference signal actually receives interference power is q_1. Similarly, there is one way to choose both adjacent channels with interference signals and $\binom{M-3}{N_s-N_1-2}$ ways to choose $N_s - N_1 - 2$ channels with interference signals out of $M - 3$ channels. The probability that exactly one of the two adjacent channels with interference signals actually receives interference power is $2q_1(1 - q_1)$. Because of the sector synchronization, either all of the signals from a sector overlap a desired symbol with probability q_2 or none of them do. Therefore, the probability that a symbol is degraded by interference in exactly one of the adjacent channels of the communicators is

$$P_{a1} = \frac{\binom{2}{1}\binom{M-3}{N_s-N_1-1}}{\binom{M-1}{N_s-N_1}} q_1 q_2 + \frac{\binom{M-3}{N_s-N_1-2}}{\binom{M-1}{N_s-N_1}} 2\, q_1(1-q_1)q_2$$

$$= \frac{2d(N_s-N_1)}{(M-1)(M-2)}[M-2-q_1(N_s-N_1-1)], \quad M \geq N_s \,. \quad (7\text{-}16)$$

Similarly, the probability that a symbol is degraded by interference in both adjacent channels is

$$P_{a2} = \frac{\binom{M-3}{N_s-N_1-2}}{\binom{M-1}{N_s-N_1}} q_1^2 q_2$$

$$= \frac{dq_1\,(N_s-N_1)\,(N_s-N_1-1)}{(M-1)(M-2)}\,, \quad M \geq N_s \,. \quad (7\text{-}17)$$

For adjacent-channel interference from within sector A, P_{a1} and P_{a2} are given by the same equations with $N_1 = 1$ to reflect the fact that one of the mobiles is the communicating mobile.

Suppose that separated orthogonal frequency-hopping patterns with $\nu = 2$ are used. There is no adjacent-channel interference from sector A. If a signal from an interfering sector B uses the transmission channel so that $N_1 = 1$, an event with probability N_s/M, then the carrier separation of the signals generated in sector B ensures that there is no adjacent-channel interference from sector B. Suppose that no signal from sector B uses the transmission channel so that $N_1 = 0$. Interference in exactly one adjacent channel results if the transmission channel of the desired signal in sector A, which may be any of $M - N_s$ channels, is located next to one of the two end channels of the set of N_s separated channels being used in sector B, neglecting hopset end effects. It also results if the transmission channel is located between two separated channels, of which only one is currently being used in sector B, again neglecting hopset end effects. Therefore, the probability that a symbol is degraded by interference in exactly one of the adjacent channels of the communicators is

$$P_{a1} = \left[\frac{2q_1}{M-N_s} + \frac{N_s-1}{M-N_s} q_1(1-q_1)\right] q_2$$

$$= \frac{d}{M-N_s}\left[(N_s-1)\,(1-q_1)+2\right]\,, \quad M \geq 2N_s\,,\; N_1 = 0\,. \quad (7\text{-}18)$$

Interference in both adjacent channels results if the transmission channel is located between two separated channels of sector B and both are being used, neglecting hopset end effects. Therefore, the probability that a symbol is degraded by interference in both adjacent channels is

$$P_{a2} = \frac{dq_1\,(N_s-1)}{M-N_s}\,, \quad M \geq 2N_s\,,\; N_1 = 0. \quad (7\text{-}19)$$

7.5 Cellular Simulation Results

In the simulation, the spatial configuration consists of a hexagonal grid of cells with base stations at their centers. Each cell has a radius R_0 from its center to a corner. A central cell is surrounded by an inner concentric tier of 6 cells and an outer concentric tier of 12 cells, as depicted in Figure 27. Other tiers are assumed to generate insignificant interference in the central cell. An equal number of mobiles, each transmitting at the same power level, is located in each sector and served by that sector's antenna. This assumption is pessimistic since slightly improved performance may be possible if a mobile is served by the sector antenna providing a signal with the least attenuation, and if hysteresis effects during handoffs are not too severe. Each signal transmitted by a sector antenna is allocated the same power. The set of frequency-hopping patterns used in each sector is assumed to be selected independently of the other sectors. Since the parameter R_0 in (7-7) and (7-9) is equal to the maximum communication range, p_0 is the minimum received area-mean power of a desired signal. The location of each mobile within a sector is assumed to be uniformly distributed.

In each simulation trial for communications in sector A of the central cell, the location of the desired mobile is randomly selected according to the uniform distribution. The selected distance of the desired mobile is substituted into (7-9) as the value of r, and then (7-9) is used to randomly select the local-mean power of the desired signal at the receiver. Each transmitting and receiving beam produced by a sector antenna is assumed to have a constant gain over its sector and zero gain elsewhere.

For an uplink of sector A, interference is assumed to arrive from mobiles within sector A, mobiles in the 6 sectors of the two cells in the inner tier that were covered by the beam of sector A, and mobiles in the 11 complete sectors and 2 half-sectors of the five cells in the outer tier completely or partially covered by the beam. The 2 half-sectors are approximated by an additional complete sector in the outer tier. Equations (7-15) to (7-19) are used to determine if a sector contains mobiles that produce power in the transmission channel or in one or both of the adjacent channels. If the sector does, then the locations of the three or fewer interfering mobiles are randomly selected according to the uniform distribution, and their distances from the central cell's base station are computed.

For a downlink of sector A, interference is assumed to arrive from the facing sector antenna of each cell in the two surrounding tiers. Equations (7-15) to (7-19) are used to determine if a signal generated by an interfering sector antenna produces power in the transmission channel or the adjacent channels of the desired signal. If so, then the distance between the sector antenna and the desired mobile is computed.

If the power from an interferer enters the transmission channel, then the power level is randomly selected according to (7-9), with the appropriate distance substituted. If the power enters an adjacent channel, then the potential local-mean power level is first randomly selected via (7-9) and then multiplied by K_s to

determine the net interference power p_{ui} that appears in (7-10). The shadowing parameter σ_s is assumed to be the same for all signals originating from all cells. The effects of p_0 and σ_n^2 are determined solely by p_0/σ_n^2, the minimum area-mean SNR. Since only ratios affect the performance, the numerical value of R_0 in the simulation is immaterial and is set equal to unity.

Once the local-mean power levels and the noise power are calculated, the symbol error probability is calculated with (7-10) and (7-11). Each simulation experiment was repeated for 20,000 trials, with different randomly selected mobile locations in each trial. The performance measure is the spatial reliability, which is a function of $\bar{\rho}$, the SINR. The appropriate value of the threshold E depends on the desired information-bit error probability and the error-correcting code.

Figures 28 to 31 depict the results of four simulation experiments for the uplinks of a cellular network. The figures plot spatial reliability as a function of N_s for various values of L, assuming MSK, three sectors, and that $\beta = 4$, $q_1 = 0.4$, $q_2 = 1.0$, $\sigma_s = 8$ dB, $E = 0.01$, $\zeta = 1$, and $K_s = 0.015$. The value of K_s results from assuming contiguous frequency channels with the center frequencies separated by B.

Figure 28 provides a baseline with which other figures may be compared. For this figure, separated orthogonal frequency hopping with $\nu = 2$, $M = 100$, and minimum area-mean SNR = 30 dB are assumed. The figure illustrates the dramatic performance improvement provided by dual spatial diversity when

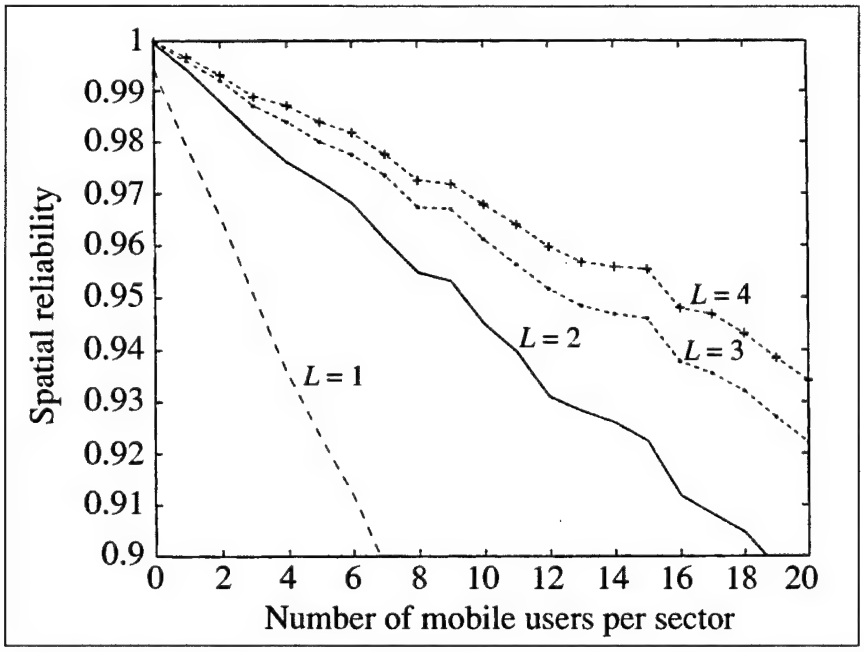


Figure 28. Spatial reliability for uplinks, separated orthogonal hopping, $M = 100$, and minimum area-mean SNR = 30 dB.

Rayleigh fading occurs. Further increases in diversity yield diminishing returns. One can assess the impact of the spectral splatter in this example by setting $K_s = 0$ and observing the change in spatial reliability. The change is insignificant because by far the most potentially damaging splatter arises from mobiles in the same sector as the desired mobile, and the separated orthogonality has eliminated it.

Figure 29 shows the effect of using orthogonal rather than separated orthogonal frequency hopping. The performance loss is significant in this example and becomes more pronounced as M decreases. When separated orthogonal frequency hopping is used and the spectral splatter is negligible, then the spatial reliability depends primarily on $M_1 = M/d$, the equivalent number of channels. In Figure 28, $M_1 = 250$.

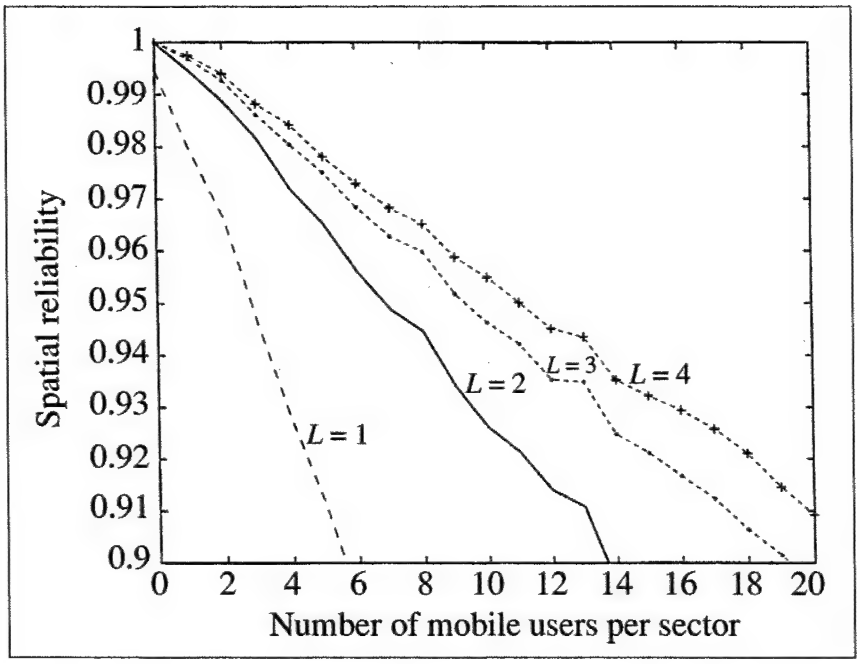


Figure 29. Spatial reliability for uplinks, orthogonal hopping, M = 100, and minimum area-mean SNR = 30 dB.

Figure 30 illustrates the effect of increasing M to 200, and hence increasing M_1 to 500. The *uplink capacity* C_u of a cellular network is defined as the maximum number of interfering mobiles per cell for which the spatial reliability exceeds 0.95. Figures 28 and 30 and other simulation results indicate that for three sectors per cell, dual diversity, and the other parameter values selected, the uplink capacity is $C_u \approx 0.108\ M_1$ for $50 \leq M_1 \leq 1000$. This equation is sensitive to parameter variations. If the shadowing standard deviation σ_s is lowered to 6 dB, it is found that C_u increases by roughly 57 percent. Alternatively, if the threshold E is raised to 0.04, corresponding to SINR = 7 dB, it is found that C_u increases by roughly 59 percent.

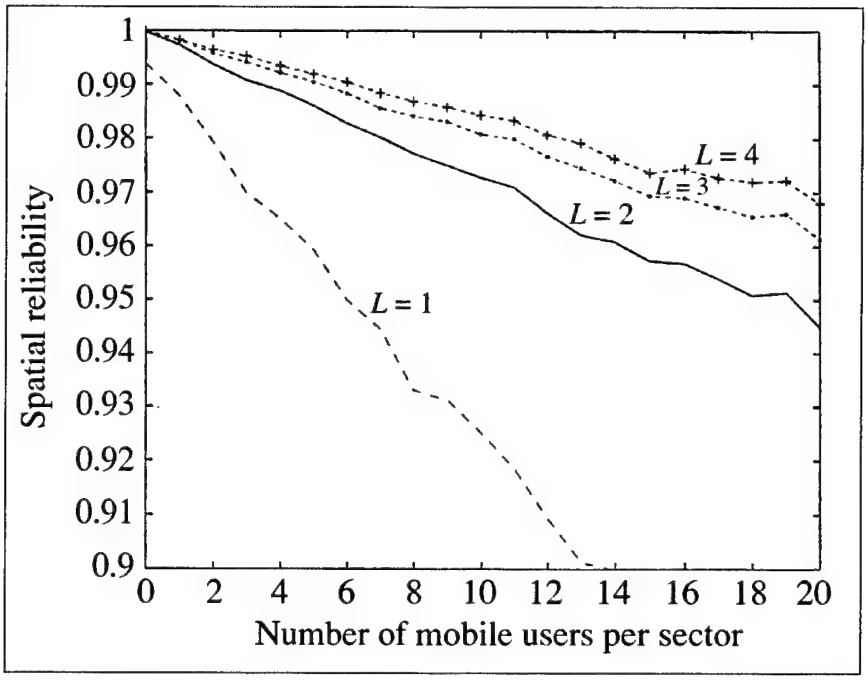


Figure 30. Spatial reliability for uplinks, separated orthogonal hopping, M = 200, and minimum area-mean SNR = 30 dB.

Figure 31 illustrates the sensitivity of the network to a decrease in the minimum area-mean SNR, which may be due to a change in either p_0 or σ_n^2. A substantial performance loss occurs when the minimum area-mean SNR is reduced to 20 dB, particularly for no spatial diversity or dual diversity. Other simulation results indicate that an increase in the minimum area-mean SNR beyond 30 dB barely improves performance.

The downlinks of a cellular network are considered in Figure 32, where the models and parameter values are otherwise the same as in Figure 28. The performance is worse for the downlinks of Figure 32 than for the uplinks of Figure 28 because of the relative proximity of some of the interfering sector antennas to the desired mobile. The *downlink capacity* C_d, which is defined analogously to the uplink capacity, is $C_d \approx 0.072\ M_1$ for $50 \leq M_1 \leq 1000$. A more realistic comparison of the downlinks and uplinks must take into account the differences between the high-power amplifiers and low-noise amplifiers in the base station and those in the mobiles. Assuming a net 10 dB advantage in the minimum area-mean SNR for the downlinks, Figures 31 and 32 provide a performance comparison of the two links. The performance of the downlinks is still slightly worse if $L \geq 2$ and $N_s \geq 4$. The difference in performance is further increased if physical constraints limit the downlinks to L = 1 or 2 while $L = 4$ for the uplinks.

Compared with direct-sequence systems, frequency-hopping systems have a bandwidth advantage in that frequency hopping over a large, possibly noncontiguous, spectral band is as practical as direct-sequence spreading over a much smaller, necessarily contiguous, spectral band. Even deprived of its

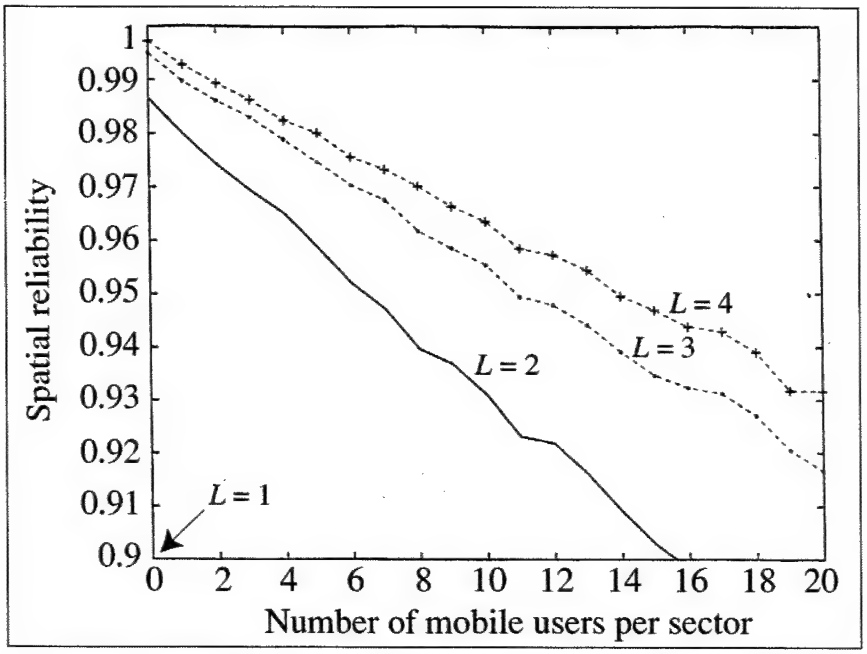


Figure 31. Spatial reliability for uplinks, separated orthogonal hopping, M = 100, and minimum area-mean SNR = 20 dB.

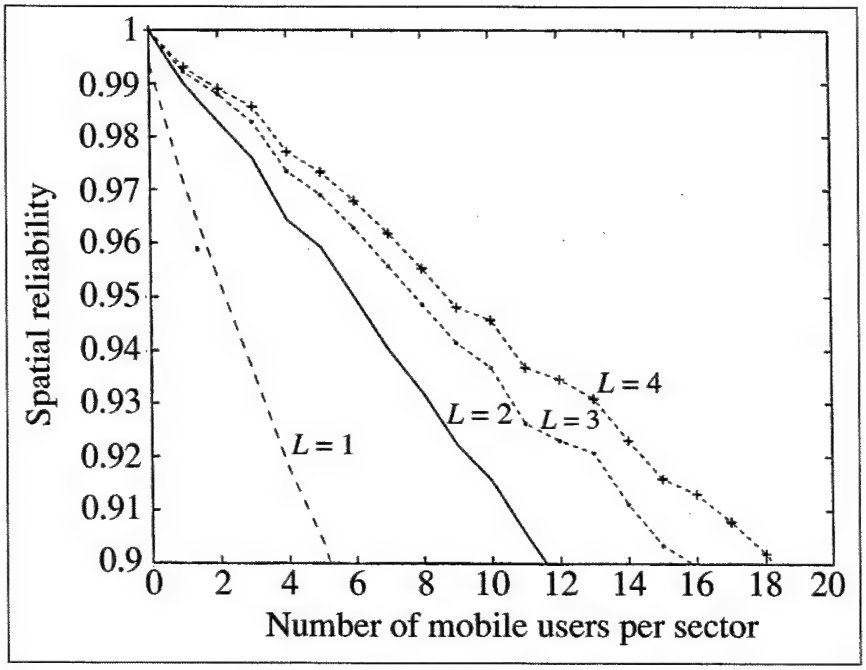


Figure 32. Spatial reliability for downlinks, separated orthogonal hopping, M = 100, and minimum area-mean SNR = 30 dB.

bandwidth advantage, as well as power control and the use of one-coincidence codes, frequency-hopping CDMA can provide nearly the same multiple-access capacity over the uplinks as direct-sequence CDMA subject to realistic power-control imperfections [22].

7.6 Summary

The performance of frequency-hopping CDMA communications in a mobile peer-to-peer network is greatly improved by the use of spatial diversity, which usually requires carrier frequencies in excess of 1 GHz. A crucial parameter is the number of equivalent frequency channels, which can be increased not only by an increase in the number of frequency channels, but also by a decrease in the duty factor of the network users. The data modulation method that is most suitable appears to be MSK or some other form of CPFSK or CPM. For these modulations, $BT_s \approx 1$, and the scenario modeled, the spectral splatter from adjacent channels, is not an important factor if the number of interferers is much smaller than the number of equivalent channels.

For a specified sectorization, diversity, and waveform, the capacity of a cellular frequency-hopping CDMA network is approximately proportional to the equivalent number of frequency channels. Thus, a desired capacity can be attained by choosing a sufficiently large number of frequency channels. A major advantage of frequency hopping is that these channels do not have to be spectrally contiguous but can be scattered throughout a large spectral band. Another advantage is that power control is not required. Its absence allows a substantial reduction of system complexity and overhead cost and facilitates geolocation. Sectorization, orthogonality, and dual diversity are invaluable, but higher levels of diversity offer sharply decreasing gains. If spectral splatter is a problem, separated orthogonal signaling can be used to eliminate it. The overall limit on the capacity of a frequency-hopping CDMA network appears to be set more by the downlinks than the uplinks.

Frequency hopping may be added to almost any communication system to strengthen it against interference or fading. For example, the set of carriers used in a multicarrier direct-sequence CDMA system or the subcarriers of an *orthogonal frequency-division-multiplexing* (OFDM) system may be hopped. The purpose of OFDM is to enable high data-rate communications without an elaborate equalization system by transmitting symbols simultaneously over a number of subcarriers.

8. Synchronization of Frequency-Hopping Patterns

The synchronization of the reference frequency-hopping pattern produced by the receiver synthesizer with the received pattern has two stages: acquisition and tracking. During *acquisition*, the reference pattern is synchronized with the received pattern to within a fraction of a hop duration. The *tracking* system further reduces the synchronization error, or at least maintains it within certain bounds. For communication systems that require a strong capability to reject interference, *matched-filter acquisition* and *serial-search acquisition* are the most effective techniques. The matched filter provides rapid acquisition of short frequency-hopping patterns, but requires the simultaneous synthesis of multiple frequencies. The matched filter may also be used to detect short patterns embedded in much longer frequency-hopping patterns. Such a detection can be used to initialize or supplement serial-search acquisition, which is more reliable and accommodates long patterns.

8.1 Matched-Filter Acquisition

Figure 33 shows a programmable *matched-filter acquisition system* that provides substantial protection against interference [27]. It is assumed that a single frequency channel is used during each hop interval that occurs during acquisition. One or more programmable frequency synthesizers produce tones at frequencies $f_1, f_2, \ldots, f_N$, which are offset by a constant frequency from the consecutive frequencies of the hopping pattern for code acquisition. Each tone multiplies the received frequency-hopping signal and the result is filtered so that most of the received energy is blocked, except the energy in a frequency-hopping pulse at a specific frequency. The threshold detector of branch k produces $d_k(t) = 1$ if its threshold is exceeded, which ideally occurs only if the received signal hops to a specific frequency. Otherwise, the threshold detector produces $d_k(t) = 0$. The use of binary detector outputs prevents the system from being overwhelmed by a few strong interference signals. Input $D(t)$ of the comparator is the number of frequencies in the hopping pattern that were received in succession. This discrete-valued, continuous-time function is

$$D(t) = \sum_{k=1}^{N} d_k[t - (N - k + 1)T_h] \tag{8-1}$$

where T_h is the hop duration. These waveforms are illustrated in Figure 34(a) for $N = 8$. The input to the threshold generator is

$$L(t) = D(t + T_h) \tag{8-2}$$

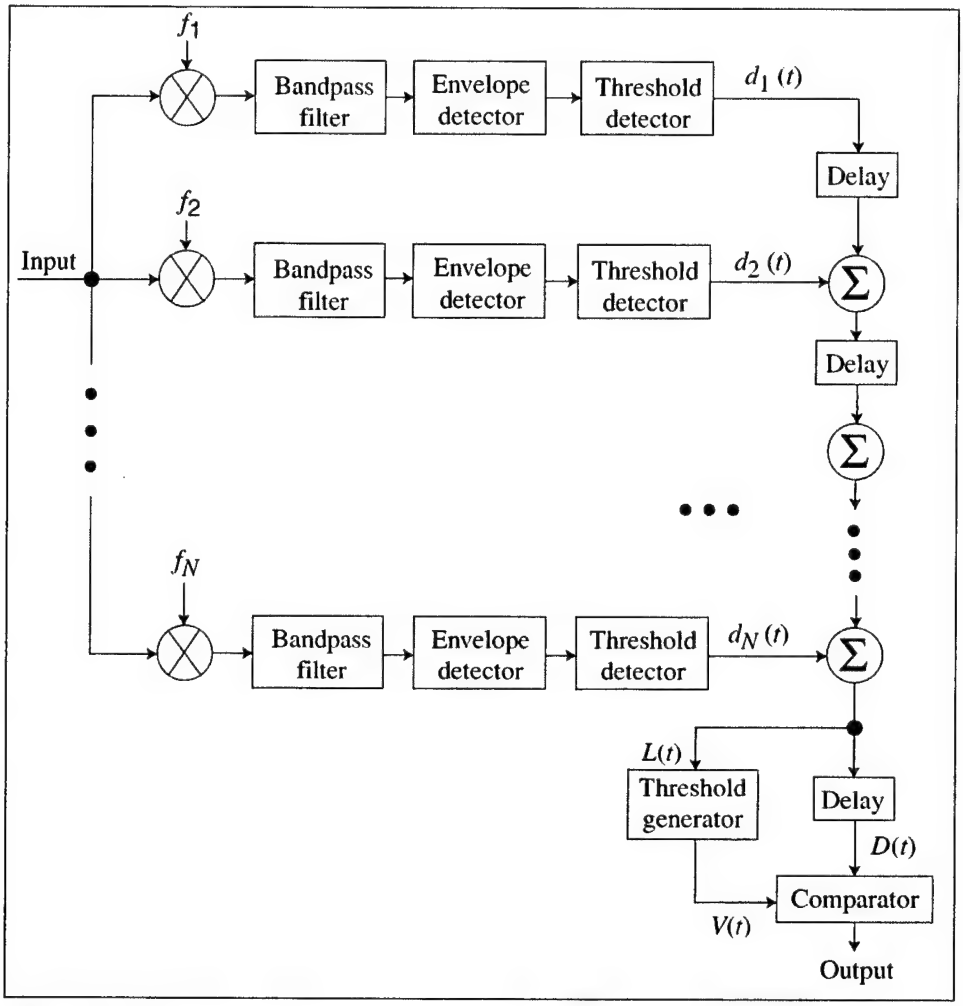


Figure 33. Matched-filter acquisition system with protection against interference.

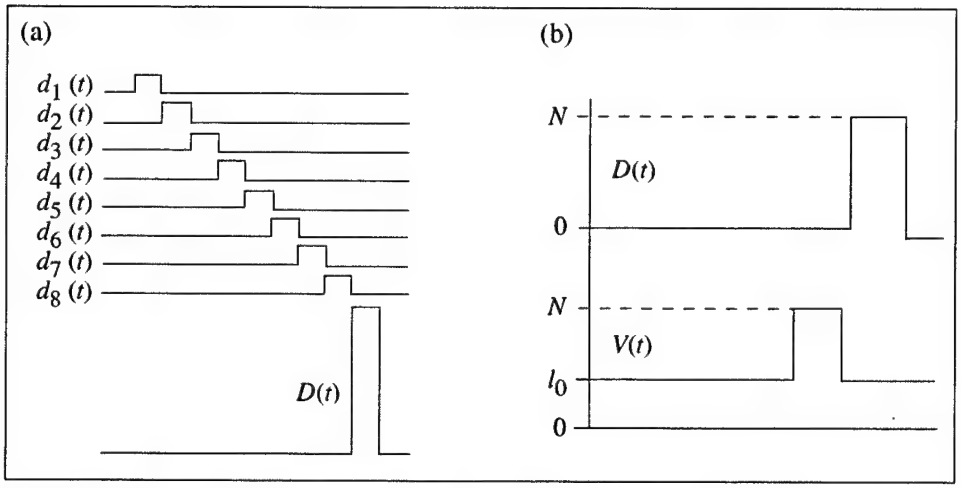


Figure 34. Ideal acquisition system waveforms: (a) formation of $D(t)$ when $N = 8$, and (b) comparison of $D(t)$ and $V(t)$.

Acquisition is declared when $D(t) \geq V(t)$, where $V(t)$ is an adaptive threshold that is a function of $L(t)$. An effective choice is

$$V(t) = \min[L(t) + l_0, N] \tag{8-3}$$

where l_0 is a positive integer. In the absence of noise and interference, $L(t) = 0$ and $V(t) = l_0$ during the hop interval in which $D(t) = N$, as illustrated in Figure 34(b). If j of the N frequency channels monitored by the matched filter receive strong, continuous interference, then $L(t) = j$ and $V(t) = j + l_0$ during this hop interval if $j \leq N - l_0$, and $D(t) \geq V(t)$. During other intervals, $j + l_0 \leq V(t) \leq N$, but $D(t) = j$. Therefore, $V(t) > D(t)$, and the matched filter does not declare acquisition. False alarms are prevented because $L(t)$ provides an estimate of the number of frequency channels with continuous interference.

When acquisition tone k is received, the signal in branch k of the matched filter is

$$r_k(t) = \sqrt{2S}\cos 2\pi f_0 t + \sqrt{2I}\cos(2\pi f_0 t + \phi) + n(t) \tag{8-4}$$

where f_0 is the intermediate frequency, the first term is the desired signal with average power S, the second term represents tone interference with average power I, $n(t)$ is zero-mean, stationary Gaussian noise and interference, and ϕ is the phase shift of the tone interference relative to the desired signal. The power in $n(t)$ is

$$N_1 = N_t + N_i \tag{8-5}$$

where N_t is power of the thermal noise and N_i is the power of the statistically independent noise interference.

Bandpass filters are used instead of filters matched to the acquisition tones because the appropriate sampling times are unknown. The passbands of the bandpass filters in the branches are assumed to be spectrally disjoint so that tone interference entering one branch has negligible effect on the other branches, and the filter outputs are statistically independent of each other. To prove the statistical independence of the noise, let $R_n(\tau)$ and $S_n(f)$ denote the autocorrelation and power spectral density, respectively, of the stationary Gaussian noise $n(t)$ in the received signal. Let $h_1(t)$ and $h_2(t)$ denote the impulse responses and $H_1(f)$ and $H_2(f)$ the transfer functions of two bandpass filters. Since the same Gaussian noise process enters both filters, their outputs are jointly Gaussian. The cross-covariance of the jointly Gaussian, zero-mean filter outputs is

$$\begin{aligned} C &= E\left[\int h_1(\tau_1)n(t-\tau_1)d\tau_1 \int h_2(\tau_2)n(t-\tau_2)d\tau_2\right] \\ &= \int\int h_1(\tau_1)h_2(\tau_2)R_n(\tau_2-\tau_1)d\tau_1\, d\tau_2 \\ &= \int\int\int h_1(\tau_1)h_2(\tau_2)S(f)\exp[j2\pi f(\tau_2-\tau_1)]df\, d\tau_1\, d\tau_2 \\ &= \int S(f)H_1(f)H_2^*(f)df \end{aligned} \tag{8-6}$$

where all the integrals extend over $(-\infty, \infty)$. Thus, $C = 0$, if $H_1(f)$ and $H_2(f)$ are spectrally disjoint. If the noise is white and, hence, $S(f)$ is a constant, then $C = 0$ if $H_1(f)$ and $H_2(f)$ are orthogonal. When $C = 0$ for all pairs of bandpass filters, the threshold-detector outputs in the N branches are statistically independent.

Suppose that noise interference is present in a branch, but that tone interference is absent so that $I = 0$. The stationary Gaussian noise has the representation (Appendix B-3)

$$n(t) = n_c(t)\cos 2\pi f_0 t - n_s(t)\sin 2\pi f_0 t \tag{8-7}$$

where $n_c(t)$ and $n_s(t)$ are zero-mean Gaussian processes with noise powers equal to N_1. In practice, the matched filter of Figure 33 would operate in continuous time so that acquisition might be declared at any moment. However, for analytical simplicity, the detection and false-alarm probabilities are calculated under the assumption that there is one sample taken per hop dwell time. From (8-5) with $I = 0$ and (8-7), it follows that

$$r_k(t) = \sqrt{Z_1^2(t) + Z_2^2(t)}\cos[2\pi f_0 t + \psi(t)] \tag{8-8}$$

where

$$Z_1(t) = \sqrt{2S} + n_c(t)\ ,\quad Z_2(t) = n_s(t)\ ,\quad \psi(t) = \tan^{-1}\left[\frac{n_s(t)}{n_c(t)}\right] \tag{8-9}$$

Since $n_c(t)$ and $n_s(t)$ are statistically independent (Appendix B.3), the joint probability density function of Z_1 and Z_2 at any specific time is

$$g_1(z_1, z_2) = \frac{1}{2\pi N_1}\exp\left[-\frac{(z_1 - \sqrt{2S})^2 + z_2^2}{2N_1}\right] \tag{8-10}$$

Let R and Θ be implicitly defined by $Z_1 = R\cos\Theta$ and $Z_2 = R\sin\Theta$. The joint density of R and Θ is

$$g_2(r, \theta) = \frac{r}{2\pi N_1}\exp\left(-\frac{r^2 - 2r\sqrt{2S}\cos\theta + 2S}{2N_1}\right),\ r \geq 0\ ,\ |\theta| \leq \pi \tag{8-11}$$

The probability density function of the envelope-detector output $R = \sqrt{Z_1^2(t) + Z_2^2(t)}$ is obtained by integration over θ. Using (A-59) gives

$$f_1(r) = \frac{r}{N_1}\exp\left(-\frac{r^2 - 2S}{2N_1}\right)I_0\left(r\frac{\sqrt{2S}}{N_1}\right)u(r) \tag{8-12}$$

where $I_0(\ )$ is the modified Bessel function of the first kind and order zero, and $u(r) = 1$ if $r \geq 0$ and $u(r) = 0$ if $r < 0$.

The detection probability for the threshold detector in the branch is

$$P_{11} = \int_\eta^\infty f_1(r)dr \tag{8-13}$$

where η is the threshold. The *Q-function* is defined as

$$Q(\alpha, \beta) = \int_{\beta}^{\infty} x \exp\left(-\frac{x^2 + \alpha^2}{2}\right) I_0(\alpha x) dx \tag{8-14}$$

Applying this definition,

$$P_{11} = Q\left(\sqrt{\frac{2S}{N_1}}, \frac{\eta}{\sqrt{N_1}}\right) \tag{8-15}$$

In the absence of noise interference, the detection probability is

$$P_{10} = Q\left(\sqrt{\frac{2S}{N_t}}, \frac{\eta}{\sqrt{N_t}}\right) \tag{8-16}$$

If the acquisition tone is absent, but the noise interference is present, the false-alarm probability is

$$P_{01} = \exp\left(-\frac{\eta^2}{2N_1}\right) \tag{8-17}$$

In the absence of both the acquisition tone and the noise interference, the false-alarm probability is

$$P_{00} = \exp\left(-\frac{\eta^2}{2N_t}\right) \tag{8-18}$$

In (8-15) to (8-18), the first subscript is 1 when the acquisition tone is present and 0 otherwise, whereas the second subscript is 1 when interference is present and 0 otherwise.

Suppose that tone interference is present in a branch. We make the pessimistic assumption that this tone has a frequency exactly equal to that of the acquisition tone, as indicted in (8-4). A trigonometric expansion of the interference term and a derivation similar to that of (8-15) indicates that given the value of ϕ, the conditional detection probability is

$$P_{11}(\phi) = Q\left(\sqrt{\frac{2(S + I + \sqrt{SI}\cos\phi)}{N_1}}, \frac{\eta}{\sqrt{N_1}}\right) \tag{8-19}$$

If ϕ is modeled as a random variable uniformly distributed over $[0, 2\pi)$, then the detection probability is

$$P_{11} = \frac{1}{\pi} \int_0^{\pi} P_{11}(\phi) d\phi \tag{8-20}$$

where the fact that $\cos\phi$ takes all its possible values over $[0, \pi]$ has been used to shorten the integration interval. If the acquisition tone is absent, but the tone interference is present, the false-alarm probability is

$$P_{01} = Q\left(\sqrt{\frac{2I}{N_1}}, \frac{\eta}{\sqrt{N_1}}\right) \tag{8-21}$$

It is convenient to define the function

$$\beta(i,N,m,P_a,P_b)=\sum_{j=0}^{i}\binom{m}{j}\binom{N-m}{i-j}P_a^j(1-P_a)^{m-j}P_b^{i-j}(1-P_b)^{N-m-i+j} \tag{8-22}$$

where $\binom{b}{a}=0$ if $a>b$. Given that m of the N matched-filter branches receive interference of equal power, let the index j represent the number of interfered channels with detector outputs above η. If $0\le j\le i$, there are $\binom{m}{j}$ ways to choose j channels out of m and $\binom{N-m}{i-j}$ ways to choose $i-j$ channels with detector outputs above η from among the $N-m$ channels that are not interfered. Therefore, the conditional probability that $D(t)=i$ given that m channels receive interference is

$$P(D=i|m)=\beta(i,N,m,P_{h1},P_{h0})\ ,\quad h=0,1 \tag{8-23}$$

where h = 1 if the acquisition tones are present and h = 0 if they are not. Similarly, given that m of N acquisition channels receive interference, the conditional probability that $L(t)=l$ is

$$P(L=l|m)=\beta(l,N,m,P_{h1},P_{h0})\ ,\quad h=0,1 \tag{8-24}$$

If there are J interference signals randomly distributed among a hopset of M frequency channels, then the probability that m out of N matched-filter branches have interference is

$$P_m=\frac{\binom{N}{m}\binom{M-N}{J-m}}{\binom{M}{J}} \tag{8-25}$$

The probability that acquisition is declared at a particular sampling time is

$$P_A=\sum_{m=0}^{\min(N,J)}P_m\sum_{l=0}^{N}P(L=l|m)\sum_{k=V(l)}^{N}P(D=k|m) \tag{8-26}$$

When the acquisition tones are received in succession, the probability of detection is determined from (8-23) to (8-26). The result is

$$P_D=\sum_{m=0}^{\min(N,J)}\frac{\binom{N}{m}\binom{M-N}{J-m}}{\binom{M}{J}}\sum_{l=0}^{N}\beta(l,N,m,P_{01},P_{00})\sum_{k=V(l)}^{N}\beta(k,N,m,P_{11},P_{10}) \tag{8-27}$$

For simplicity in evaluating the probability of a false alarm, we ignore the sampling time preceding the peak value of $D(t)$ in Figure 34 because this

probability is negligible at that time. Since the acquisition tones are absent, the probability of a false alarm is

$$P_F = \sum_{m=0}^{\min(N,J)} \frac{\binom{N}{m}\binom{M-N}{J-m}}{\binom{M}{J}} \sum_{l=0}^{N} \beta(l, N, m, P_{01}, P_{00}) \sum_{k=V(l)}^{N} \beta(k, N, m, P_{01}, P_{00}) \tag{8-28}$$

If there is no interference so that $J = 0$, then (8-27) and (8-28) reduce to

$$P_D = \sum_{l=0}^{N} \binom{N}{l} P_{00}^{l}(1 - P_{00})^{N-1} \sum_{k=V(l)}^{N} \binom{N}{k} P_{10}^{k}(1 - P_{10})^{N-k} \tag{8-29}$$

$$P_F = \sum_{l=0}^{N} \binom{N}{l} P_{00}^{l}(1 - P_{00})^{N-1} \sum_{k=V(l)}^{N} \binom{N}{k} P_{00}^{k}(1 - P_{00})^{N-k} \tag{8-30}$$

The channel threshold η is selected to maintain a required P_F when there is no interference and the values of l_0, N, and N_t are given. The value of l_0 is then selected to maximize P_D given the values of N and S/N_t. The best choice is generally $l_0 = \lfloor N/2 \rfloor$. For example, suppose that $N = 8$, $P_F = 10^{-7}$, and the signal-to-noise ratio is $S/N_t = 10$ dB when an acquisition tone is received. A numerical evaluation of (8-30) then yields $\eta/\sqrt{N_t} = 3.1856$ and $l_0 = 4$ as the parameter values that maintain $P_F = 10^{-7}$ while maximizing P_D in the absence of interference. The threshold pair $\eta/\sqrt{N_t} = 3.1896$, $l_0 = 4$ is the choice when a fixed comparator threshold $V(t) = l_0$ is used instead of the adaptive threshold of (8-3). If $D(t)$ and $L(t)$ are sampled once every hop dwell interval, then the false-alarm rate is P_F/T_h.

As an example, suppose that noise jamming with total power N_{it} is uniformly distributed over J matched-filter frequency channels so that

$$N_i = \frac{N_{it}}{J} \tag{8-31}$$

is the power in each of these channels. Interference tones are absent and $N = 8$, $M = 128$, and $S/N_t = 10$ dB. To ensure that $P_F = 10^{-7}$ in the absence of jamming, we assume that $l_0 = 4$ and $\eta/\sqrt{N_t} = 3.1856$ when an adaptive comparator threshold is used, and that $l_0 = 4$ and $\eta/\sqrt{N_t} = 3.1896$ when a fixed comparator threshold is used. Since P_D is relatively insensitive to J, its effect is assessed by examining P_F. Figure 35 depicts P_F as a function of N_{it}/S, the jamming-to-signal ratio. The figure indicates that an adaptive threshold is much more resistant to partial-band jamming than a fixed threshold when N_{it}/S is large. When $N_{it}/S < 10$ dB, the worst-case partial-band jamming causes a considerably higher P_F than full-band jamming. It is found that multitone jamming tends to produce fewer false alarms than noise jamming. Various other performance and design issues and the impact of frequency-hopping interference are addressed in [27].

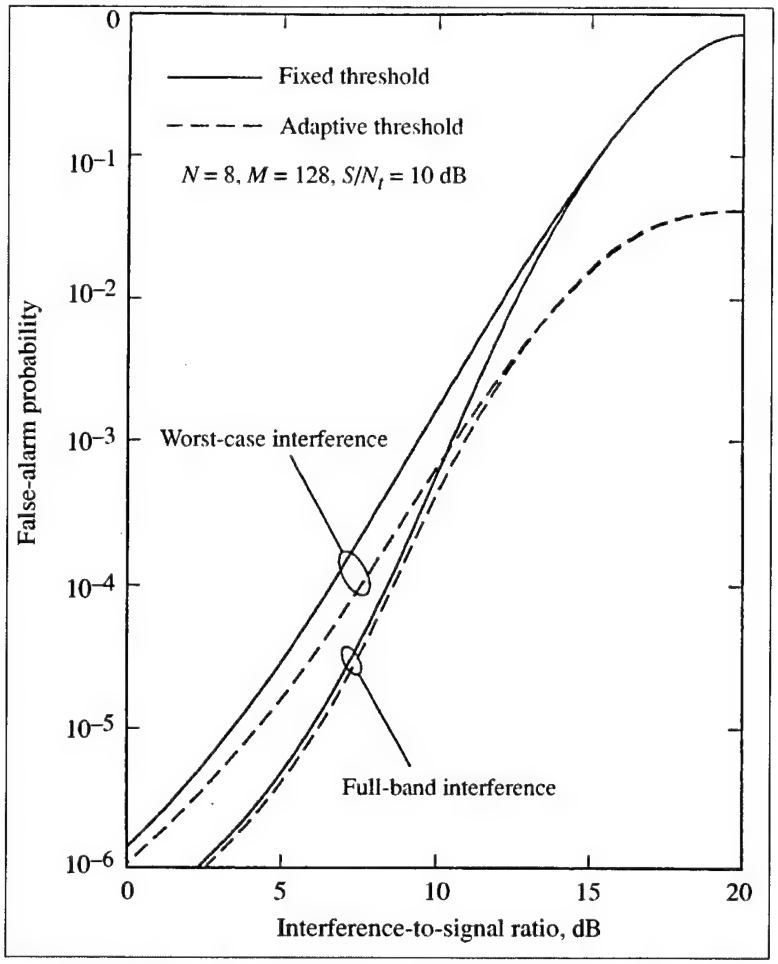


Figure 35. False-alarm probability for matched-filter acquisition system.

8.2 Serial-Search Acquisition

As illustrated by Figure 36, a *serial-search acquisition system* for frequency-hopping signals determines acquisition by attempting to develop the received frequency-hopping pattern to an intermediate frequency, and then comparing the output of an energy detector (Section 10.2) to a threshold.

If the detector integration is over several hop intervals, strong interference or deep fading over a single hop interval can cause a false alarm with high probability. This problem is mitigated by making a hard decision after integrating over each hop interval or even less in the presence of fast fading. After N decisions, a test for acquisition is passed or failed if the comparator threshold has been exceeded l_0 or more times out of N.

A trial alignment of the frequency-hopping pattern synthesized by the receiver with the received pattern is called a *cell*. If a cell passes certain tests, acquisition is declared and the tracking system is activated. If not, the cell is rejected, and a new cell is produced and tested. Let P_{Dp} and P_{Da} denote the probabilities that the comparator threshold is exceeded at the end of a hop interval when the

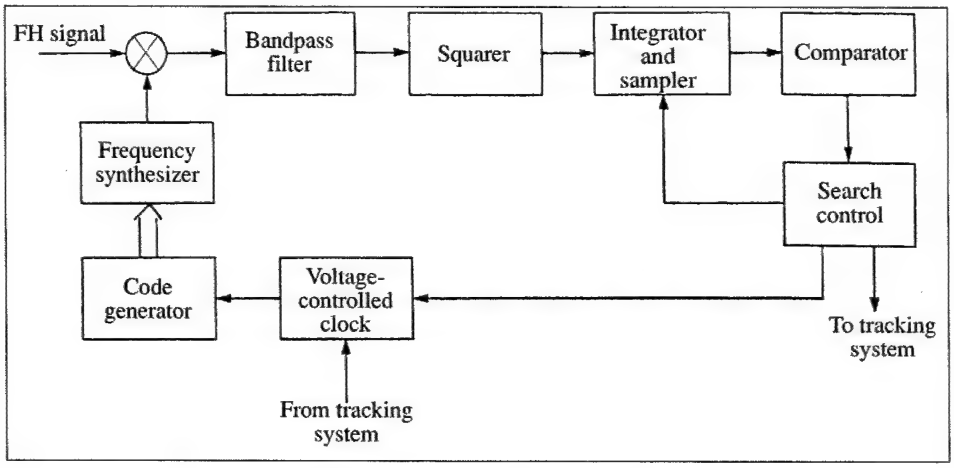


Figure 36. Serial-search acquisition system.

correct cell is tested and interference is present and absent, respectively. Let P_D denote the probability that an acquisition test is passed when the correct cell is tested. If the N acquisition tones in a test are distinct, then a derivation similar to that of Section 8.1 yields

$$P_D = \sum_{m=0}^{\min(N,J)} \frac{\binom{N}{m}\binom{M-N}{J-m}}{\binom{M}{J}} \sum_{l=l_0}^{N} \beta(l, N, m, P_{Dp}, P_{Da}) \tag{8-32}$$

where $l_0 \geq 0$. Similarly, the probability that an acquisition test is passed when an incorrect cell is tested and no acquistion tones are present is

$$P_F = \sum_{m=0}^{\min(N,J)} \frac{\binom{N}{m}\binom{M-N}{J-m}}{\binom{M}{J}} \sum_{l=l_0}^{N} \beta(l, N, m, P_{Fp}, P_{Fa}) \tag{8-33}$$

where P_{Fp} and P_{Fa} are the probabilities that the threshold is exceeded when an incorrect cell is tested and interference is present and absent, respectively. A suitable choice for l_0 is $\lfloor N/2 \rfloor$. Since the serial-search system of Figure 36 has an embedded radiometer, the performance analysis of the radiometer given in Section 10.2 can be used to obtain expressions for P_{Dp} and P_{Da}, P_{Fp}, and P_{Fa}.

To test a new cell, the reference pattern synthesized by the receiver is either advanced or delayed relative to the received pattern. A number of search techniques are illustrated in Figure 37, which depicts successive frequencies in the received pattern and six possible receiver-generated patterns. The small arrows indicate test times while the large arrows indicate times at which acquisition is declared or subsequent verification testing begins. The step size, which is the separation in hop durations between cells, is denoted by δ. Techniques (a), (b), and (c) entail inhibiting the code-generator clock after each unsuccessful test. Technique (d) advances the reference pattern by skipping frequencies in the pattern until acquisition is declared. The *small misalignment technique* (e) is effective when there is a high probability that the reference and

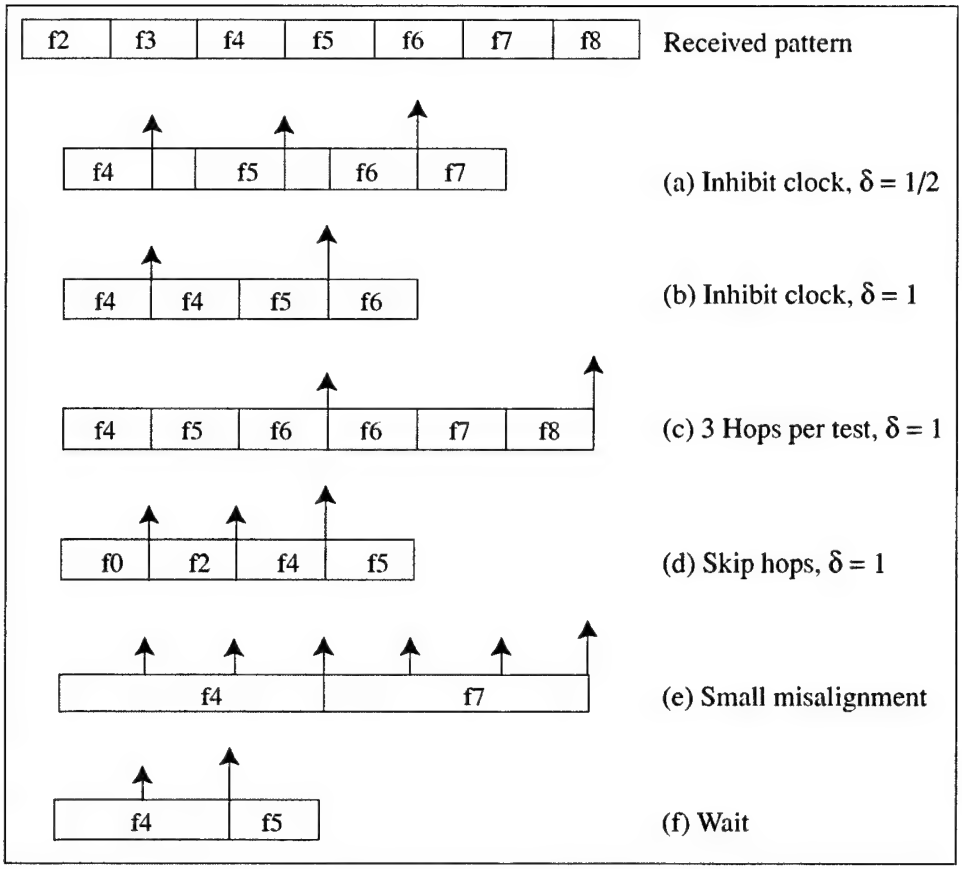


Figure 37. Search techniques for acquisition.

received patterns are within r hops of each other. The code generator temporarily forces the reference signal to remain at a frequency for $2r + 1$ hop intervals extending both before and after the interval in which the frequency would ordinarily be synthesized. If the misalignment is less than r hops, then acquisition occurs within $2r + 1$ hop durations. In the figure, $r = 1$, the initial misalignment is one-half hop duration, and it is assumed that the first time the reference and received frequencies coincide, detection fails, but the second time results in acquisition. Technique (f) entails waiting at a fixed reference frequency until this frequency is received. This technique results in a rapid search if the reference frequency is slightly advanced relative to the received pattern. The reference frequency, which may be determined by the key and TOD bits (Section 1), must be periodically shifted by at least the coherence bandwidth so that neither fading nor interference in any particular frequency channel prevents acquisition.

A reduced hopset with a short pattern period may be used temporarily to reduce the time uncertainty region and, hence, the acquisition time. In a network, a separate communication channel or cueing frequency may provide the TOD to subscribers. After detection of the TOD, a receiver might use the small misalignment technique for acquisition.

The *search control system* determines the integration intervals, the thresholds, and the logic of the tests to be conducted before acquisition is declared and the tracking system is activated. The details of the search control strategy determine the statistics of the acquisition time [8], [9], [24]. The control strategy is usually a *multiple-dwell strategy* that uses an initial test to quickly eliminate improbable cells. Subsequent tests are used for verification testing of cells that pass the initial test. The multiple-dwell strategy may be a *consecutive-count strategy*, in which a failed test causes a cell to be immediately rejected, or an *up-down strategy*, in which a failed test causes a repetition of a previous test. Figure 38 depicts the flowgraph of an up-down strategy that requires L test to be passed before acquisition is declared. The up-down strategy is preferable when the interference or noise level is high [28].

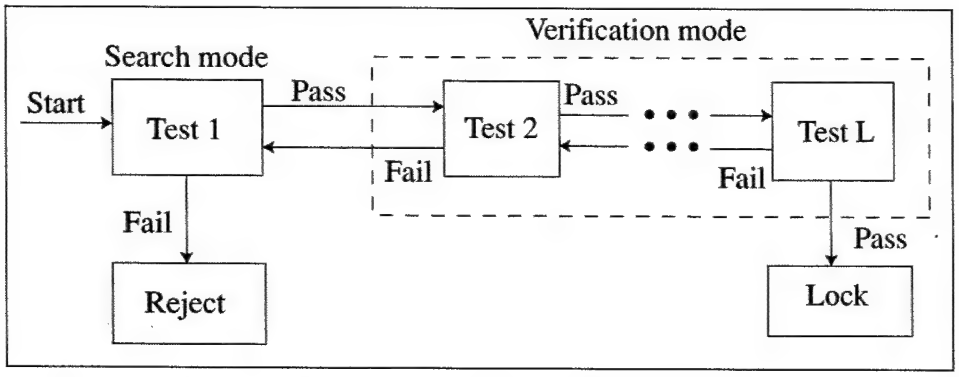


Figure 38. Flowgraph of up-down control strategy.

A verification test could be based on the comparison of the detected energy level to a threshold, as is done during the initial test. A more elaborate but potentially far more powerful procedure is to base the verification on detected bit patterns that are compared with a stored reference word. Bit synchronization may be established by oversampling the detected bits with N samples per bit and then comparing sets of samples separated by N samples with the reference word.

A large step size decreases the number of incorrect cells that must be tested before the correct cell is tested. However, as the step size increases, there is a decrease in the signal energy in the integrator output when a correct cell is tested. This issue is illustrated by Figure 39, which depicts the idealized amplitude function of the comparator input in Figure 36 for a single pulse of the received and reference signals in the absence of noise. Let τ_e denote the delay of the reference pattern relative to the received pattern. Suppose that one tested cell has $\tau_e = -x$, where $0 \le x \le \delta T_h$, and the next tested cell has $\tau_e = \delta T_h - x$ following a cell rejection. The largest amplitude of the integrator output occurs when $|\tau_e| = y$, where

$$y = \min(x, \delta T_h - x) \ , \quad 0 \le x < \delta T_h \tag{8-34}$$

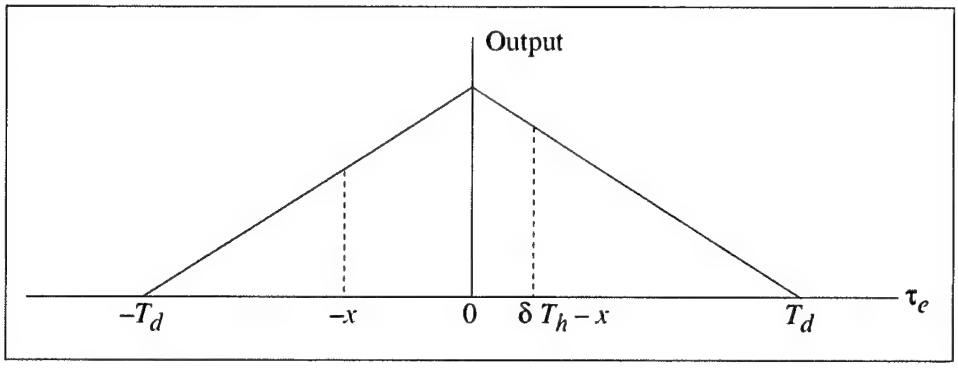


Figure 39. Amplitude of comparator input as function of relative pattern delay.

Assuming that x is uniformly distributed over $(0, \delta T_h)$, y is uniformly distributed over $(0, \delta T_h/2)$. Therefore,

$$E[y] = \frac{\delta T_h}{4} \tag{8-35}$$

$$E[y^2] = \frac{\delta^2 T_h^2}{12} \tag{8-36}$$

The *correct cell* is considered to be the one for which $|\tau_e| = y$. If the amplitude function approximates the triangular shape depicted in the figure, the comparator input when $|\tau_e| = y$ has the amplitude

$$A = A_{\max}\left(1 - \frac{y}{T_d}\right) \tag{8-37}$$

Therefore, the signal energy entering the comparator is reduced on the average by the factor

$$E\left[\left(1 - \frac{y}{T_d}\right)^2\right] = 1 - \frac{\delta T_h}{2T_d} + \frac{\delta^2 T_h^2}{12T_d^2} \tag{8-38}$$

because of the misalignment of patterns when the correct cell is tested. For example, (8-38) indicates that the average loss of 1.1 dB when $T_h = T_d$ and $\delta = 1/2$; if $\delta = 1$, then the loss is 2.3 dB.

The serial-search acquisition of frequency-hopping signals is faster than the acquisition of direct-sequence signals because the hop duration is much greater than a spreading-sequence chip duration for practical systems. Given the same time uncertainty, fewer cells have to be searched to acquire frequency-hopping signals because the step sizes are longer in time.

8.3 Tracking System

The acquisition system ensures that the receiver-synthesized frequency-hopping pattern is aligned in time with the received pattern to within a fraction of a hop duration. The tracking system must provide a fine synchronization by reducing

the residual misalignment after acquisition. The delay-locked and tau-dither loops used for the tracking of direct-sequence signals can be adapted to frequency-hopping signals [8]. However, the predominant form of tracking used in frequency-hopping systems is provided by the *early-late gate tracking loop* [29], which is shown in Figure 40 along with the ideal associated waveforms for a typical example. The course control is provided by the acquisition system. In the absence of noise, the envelope detector produces a positive output only when the received frequency-hopping signal $r(t)$, and the receiver-generated frequency-hopping replica, $r_1(t)$, are offset by the intermediate frequency f_0. The *gating signal* $g(t)$ is a square-wave clock signal with transitions from -1 to +1 that control the frequency transitions of $r_1(t)$. The early-late gate functions as a signal multiplier. Its output, $u(t)$, is the product of the gating signal and the envelope-detector output, $v(t)$. The error signal is the time integral of $u(t)$ and is a function of τ_e, the delay of $r_1(t)$ relative to $r(t)$. The error signal can be expressed as a discriminator characteristic, $e(\delta)$, which is a function of $\delta = \tau_e/T_h$, the normalized delay error. For the typical waveforms shown, δ is positive, and hence so is $e(\delta)$. Therefore, the voltage-controlled clock (VCC) will increase the transition rate of the gating signal, which will bring $r_1(t)$ into better time-alignment with $r(t)$.

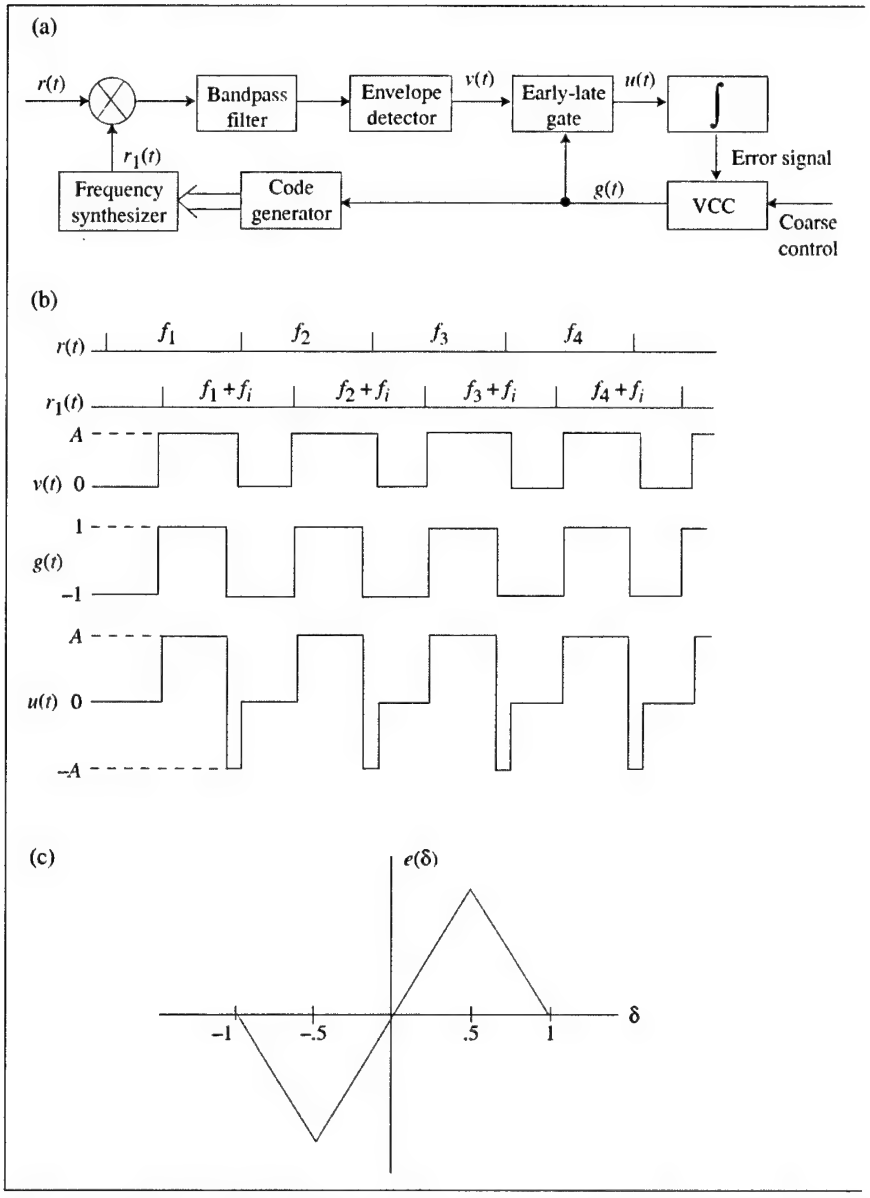


Figure 40. Early-late gate tracking: (a) loop, (b) signals, and (c) discriminator characteristic.

9. Comparison and Combination of Frequency-Hopping and Direct-Sequence Systems

Frequency-hopping systems reject interference by avoiding it, whereas direct-sequence systems reject interference by spreading it. However, the effect of partial-band interference on a frequency-hopping system is similar to the effect of high-power pulsed interference on a direct-sequence system. The interleaving and error-correcting codes that are effective against one of these types of interference are effective against the other. Error-correcting codes are more essential for frequency-hopping systems than for direct-sequence systems because partial-band interference is a more pervasive threat than high-power pulsed interference.

When frequency-hopping systems and direct-sequence systems are constrained to use the same band with a fixed bandwidth, then direct-sequence systems have an inherent advantage because they can use coherent PSK rather than a noncoherent modulation. Coherent PSK has an approximately 4 dB advantage over noncoherent MSK over the AWGN channel and an even larger advantage over fading channels. However, the potential performance advantage of direct-sequence systems is often illusory for practical reasons. A major advantage of frequency-hopping systems relative to direct-sequence systems is that it is possible to hop in frequency over a much wider band than can be occupied by a direct-sequence signal. This advantage more than compensates for the relatively inefficient noncoherent demodulation that is almost always required for frequency-hopping systems. Other major advantages of frequency hopping are its reduced susceptibility to the near-far problem, its relatively rapid acquisition, and the possibility of excluding frequency channels with steady or frequent interference.

Interleaving of the code symbols over many dwell intervals provides a large level of diversity to slow frequency-hopping systems operating over a frequency-selective fading channel. These systems are usually insensitive to variations in the Doppler spread of the channel because the additional diversity due to the time-selective fading is insignificant. In contrast, the diversity combining in a direct-sequence receiver depends on the small number of branches or demodulators in its rake receiver, and the system is relatively sensitive to variations in the Doppler spread. Slow frequency-hopping systems usually outperform comparable direct-sequence systems unless the Doppler spread is large or most of the power in the received signal is concentrated in a small number of resolvable multipath components [30].

A *hybrid frequency-hopping direct-sequence system* is a frequency-hopping system that uses direct-sequence spreading during each dwell interval or, equivalently, a direct-sequence system in which the carrier frequency changes periodically. In the transmitter of the hybrid system of Figure 41, a single code generator controls both the spreading and the hopping pattern. The spreading sequence is added modulo-2 to the data sequence. Hops occur periodically after a fixed number of sequence chips. In the receiver, the frequency hopping and the spreading sequence are removed in succession to produce a carrier with the message modulation. Because of the phase changes due to the frequency hopping, noncoherent modulation, such as DPSK, is usually required unless the hopping rate is very low. Serial-search acquisition occurs in two stages. The first stage provides alignment of the hopping patterns, whereas the second stage over the phase of the pseudonoise sequence finishes acquisition rapidly because the time uncertainty has been reduced by the first stage to less than a hop duration.

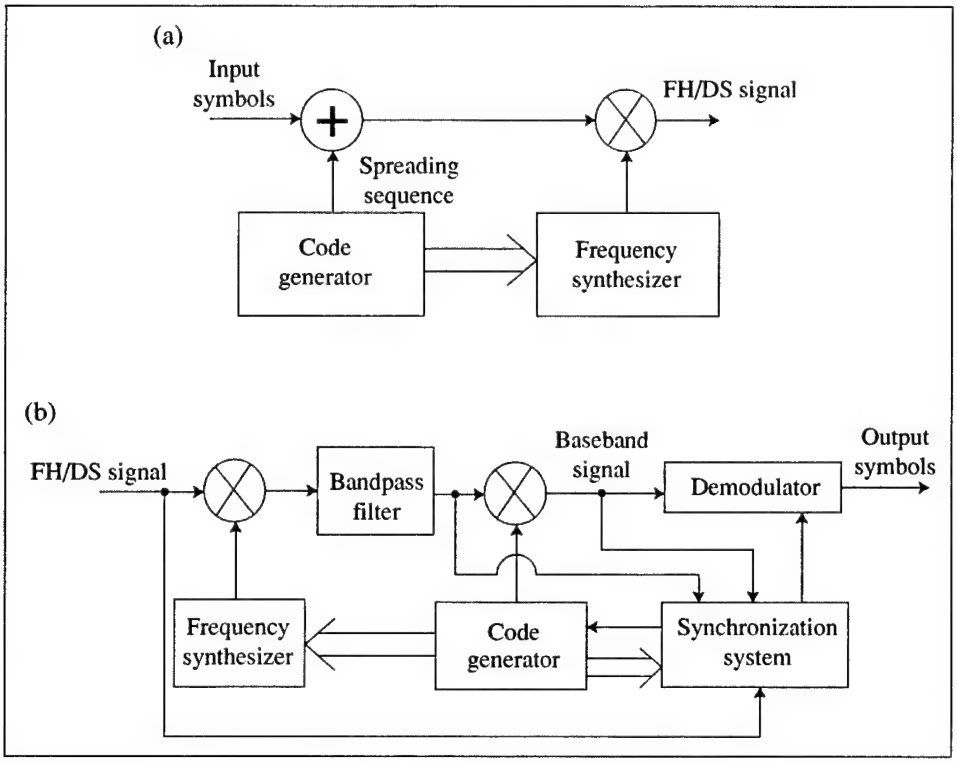


Figure 41. Hybrid frequency-hopping direct-sequence system: (a) transmitter and (b) receiver.

A hybrid system combats partial-band interference in two ways. The hopping allows the avoidance of the interference spectrum part of the time. When the system hops into the interference, the interference is spread and filtered as in a direct-sequence system. However, during a hop interval, interference that would be avoided by an ordinary frequency-hopping receiver is passed by the bandpass filter of a hybrid receiver because the bandwidth must be large enough to

accommodate the direct-sequence signal that remains after the dehopping. This large bandwidth also limits the number of available frequency channels, which increases the susceptibility to narrowband interference and the near-far problem. Thus, hybrid systems are seldom used except in specialized military applications because the additional direct-sequence spreading weakens the major strengths of frequency hopping.

10. Detection of Frequency-Hopping Signals

An interception receiver intended for the detection of frequency-hopping signals may be designed according to the principles of classical detection theory or according to more intuitive ideas. The former approach is useful in setting limits on what is possible, but the latter approach is more practical and flexible and less dependent on knowledge of the characteristics of the frequency-hopping signals.

10.1 Ideal Detection

To determine whether a signal $s(t)$ has been received over the AWGN channel based on the observation of the received signal $r(t)$, classical detection theory requires that one choose between the two hypotheses. Hypothesis H_1 is that the signal is present, whereas hypothesis H_0 is that the signal is absent. Over the observation interval $0 \leq t \leq T$, $r(t)$ under the two hypotheses is

$$r(t) = \begin{cases} s(t) + n(t) \ , & H_1 \\ n(t) \ , & H_0 \end{cases} \tag{10-1}$$

where $n(t)$ is white Gaussian noise with two-sided noise-power spectral density $N_0/2$. To enable a tractable analysis, the idealized assumptions are made that the hopset is known and that the hop epoch timing, which includes the hop-transition times and the hop rate, is known. Consider slow frequency-hopping signals with CPM (FH/CPM), which includes continuous-phase MFSK. The signal over the ith hop interval is

$$s(t) = \sqrt{2S}\cos\left[2\pi f_j t + \phi(\mathbf{d}_n, t) + \phi_i\right] \ , \ (i-1)T_h \leq t < iT_h \tag{10-2}$$

where S is the average signal power, f_j is the carrier frequency, $\phi(\mathbf{d}_n, t)$ is the CPM component that depends on the data sequence $\mathbf{d}_n$, and ϕ_i is the phase associated with the ith hop. The parameters f_j, ϕ_i, and the components of $\mathbf{d}_n$ are modeled as random variables.

The coefficients in the expansion of the observed waveform in terms of orthonormal basis functions constitute the received vector $\mathbf{r} = [r_1\, r_2\, \ldots\, r_N]$. Let $\boldsymbol{\theta}$ denote the vector of parameter values that characterize the signal to be detected. The *average likelihood ratio* [31], which is compared with a threshold for a detection decision, is

$$\Lambda(\mathbf{r}) = \frac{E_{\boldsymbol{\theta}}[f(\mathbf{r}|H_1, \boldsymbol{\theta})]}{f(\mathbf{r}|H_0)} \tag{10-3}$$

where $f(\mathbf{r}|H_1, \boldsymbol{\theta})$ is the conditional density function of $\mathbf{r}$ given hypothesis H_1 and the value of $\boldsymbol{\theta}$, $f(\mathbf{r}|H_0)$ is the conditional density function of $\mathbf{r}$ given hypothesis H_0, and $E_{\boldsymbol{\theta}}$ is the expectation over the random vector $\boldsymbol{\theta}$. The coefficients in the expansion of $n(t)$ in terms of the orthonormal basis functions are statistically independent. Since each coefficient is Gaussian with variance $N_0/2$,

$$f(\mathbf{r}|H_1, \boldsymbol{\theta}) = \prod_{i=1}^{N} \frac{1}{\sqrt{\pi N_0}} \exp\left[-\frac{(r_i - s_i)^2}{N_0}\right] \tag{10-4}$$

$$f(\mathbf{r}|H_0) = \prod_{i=1}^{N} \frac{1}{\sqrt{\pi N_0}} \exp\left(-\frac{r_i^2}{N_0}\right) \tag{10-5}$$

where the $\{s_i\}$ are the coefficients of the signal. Substituting these equations into (10-3) yields

$$\Lambda(\mathbf{r}) = E_{\boldsymbol{\theta}} \left\{ \exp\left[\frac{2}{N_0} \sum_{i=1}^{N} r_i s_i - \frac{1}{N_0} \sum_{i=1}^{N} s_i^2\right] \right\} \tag{10-6}$$

Expansions in the orthonormal basis functions indicate that if $N \to \infty$, the average likelihood ratio may be expressed in terms of the signal waveforms as

$$\Lambda[r(t)] = E_{\theta} \left\{ \exp\left[\frac{2}{N_0} \int_0^T r(t)s(t)dt - \frac{\mathcal{E}}{N_0}\right] \right\} \tag{10-7}$$

where $\mathcal{E}$ is the energy in the signal waveform over the observation interval of duration T.

The M carrier frequencies in the hopset are assumed to be equally likely over a given hop and statistically independent from hop to hop for N_h hops. Dividing the integration interval in (10-7) into N_h parts, averaging over the M frequencies, and dropping the irrelevant factor $1/M$, we obtain

$$\Lambda[r(t)] = \prod_{i=1}^{N_h} \sum_{j=1}^{M} \Lambda_{ij}[r(t)|f_j] \tag{10-8}$$

$$\Lambda_{ij}[r(t)|f_j] = E_{\boldsymbol{d}_n \phi_i} \left\{ \exp\left[\frac{2}{N_0} \int_{(i-1)T_h}^{iT_h} r(t)s(t)dt - \frac{\mathcal{E}_h}{N_0}\right] \right\} \tag{10-9}$$

where the condition in the argument of $\Lambda_{ij}[\,]$ indicates that the carrier frequency over the ith hop is f_j, the expectation is over the remaining random parameters $\boldsymbol{d}_n$ and ϕ_i, and $\mathcal{E}_h$ is the energy per hop. The decomposition in (10-8) indicates that the general structure of the detector has the form illustrated in Figure 42. The average likelihood ratio of (10-8) is compared with a threshold to determine whether a signal is present. The threshold may be set to ensure the tolerable

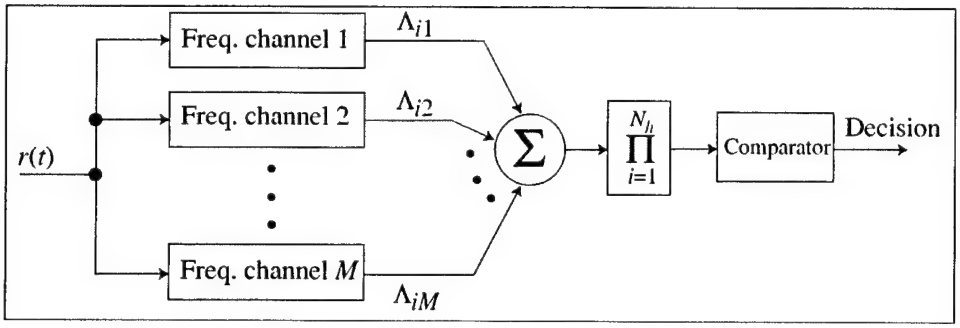


Figure 42. General structure of optimum detector for frequency-hopping signal with N_h hops and M frequency channels.

false-alarm probability when the signal is absent. Assuming that $\mathcal{E}_h = ST_h$ is the same for every hop and carrier frequency, we may drop the irrelevant factor $\exp(-\mathcal{E}_h/N_0)$ in (10-9), which only affects the threshold level.

Each of the N_d data sequences that can occur during a hop is assumed to be equally likely. For *coherent detection* of FH/CPM [32], we set $\phi_i = 0$ in (10-2), substitute it into (10-9), and then evaluate the expectation to obtain

$$\Lambda_{ij}[r(t)|f_j] = \sum_{n=1}^{N_d} \exp\left\{\frac{2\sqrt{2S}}{N_0}\int_{(i-1)T_h}^{iT_h} r(t)\cos\left[2\pi f_j t + \phi(\mathbf{d}_n, t)\right]\right\} \quad \text{(coherent)} \tag{10-10}$$

where irrelevant factors have been dropped. This equation indicates how Λ_{ij} in Figure 41 is to be calculated for each hop i and each frequency channel j corresponding to carrier frequency j. Equations (10-8) and (10-10) define the optimum coherent detector for any slow frequency-hopping signal with CPM.

For noncoherent detection of FH/CPM [32], the received carrier phase ϕ_i is assumed to be uniformly distributed over $[0, 2\pi)$ during a given hop and statistically independent from hop to hop. Averaging over the random phase in addition to the sequence statistics and dropping irrelevant factors yields

$$\Lambda_{ij}[r(t)|f_j] = \sum_{n=1}^{N_d} I_0\left(\frac{2\sqrt{2SR_{ijn}}}{N_0}\right) \quad \text{(noncoherent)} \tag{10-11}$$

where $I_0(\ )$ is defined by (A-59),

$$R_{ijn} = \left\{\int_{(i-1)T_h}^{iT_h} r(t)\cos\left[\chi_{jn}(t)\right]dt\right\}^2 + \left\{\int_{(i-1)T_h}^{iT_h} r(t)\sin\left[\chi_{jn}(t)\right]dt\right\}^2 \tag{10-12}$$

and

$$\chi_{jn}(t) = 2\pi f_j t + \phi(\mathbf{d}_n, t) \tag{10-13}$$

Equations (10-8), (10-11), (10-12), and (10-13) define the optimum noncoherent detector for any slow frequency-hopping signal with CPM. The means of producing (10-11) is diagrammed in Figure 43.

A major contributor to the huge computational complexity of the optimum detectors is the fact that with N_s data symbols per hop and an alphabet size q, there may be $N_d = q^{N_s}$ data sequences per hop. Consequently, the computational burden grows exponentially with N_s. However, if it is known that the data modulation is CPFSK with a modulation index $h = 1/K$, where K is a positive integer, the computational burden has a linear dependence on N_s [32]. Even then, the optimum detectors are extremely complex when the number of frequency channels is large.

The preceding theory may be adapted to the detection of fast frequency-hopping signals with MFSK as the data modulation. Since there is one hop per MFSK channel symbol, the information is embedded in the sequence of carrier frequences. Thus, we may set $N_d = 1$ and $\phi(\mathbf{d}_n, t) = 0$ in (10-10) and (10-11). For coherent detection, (10-10) reduces to

$$\Lambda_{ij}[r(t)|f_j] = \exp\left[\frac{2\sqrt{2S}}{N_0}\int_{(i-1)T_h}^{iT_h} r(t)\cos(2\pi f_j t)\,dt\right] \qquad \text{(coherent)} \tag{10-14}$$

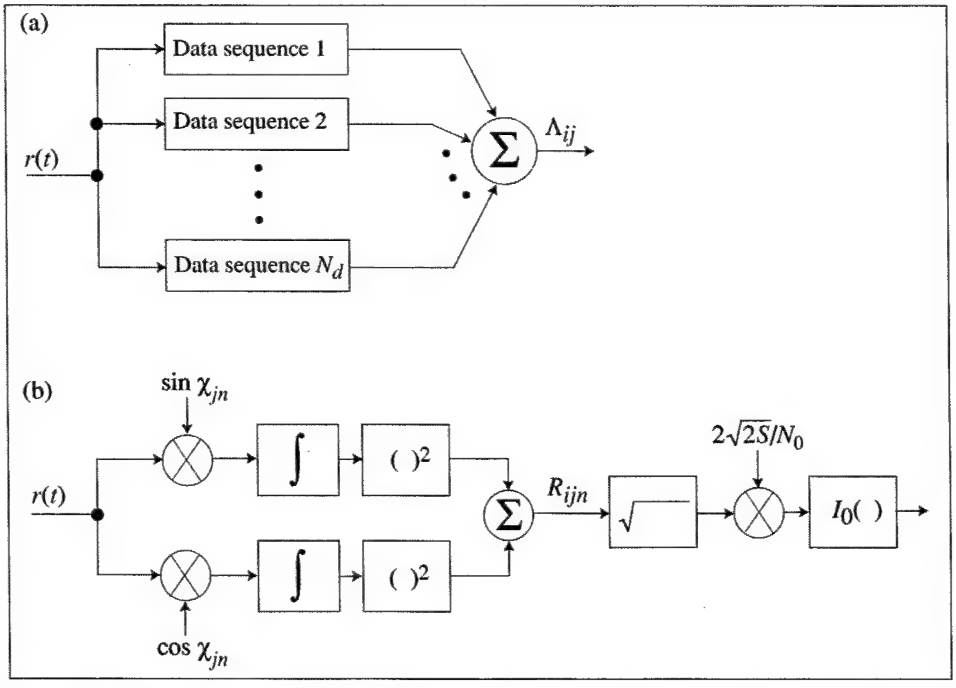


Figure 43. Optimum noncoherent detector for slow frequency hopping with CPM: (a) basic structure of frequency channel j for hop i with parallel cells for N_d candidate data sequences, and (b) cell for data sequence n.

Equations (10-8) and (10-14) define the optimum coherent detector for a fast frequency-hopping signal with MFSK. For noncoherent detection, (10-11), (10-12), and (10-13) reduce to

$$\Lambda_{ij}[r(t)|f_j] = I_0\left(\frac{2\sqrt{2SR_{ij}}}{N_0}\right) \qquad \text{(noncoherent)} \qquad (10\text{-}15)$$

$$R_{ij} = \left[\int_{(i-1)T_h}^{iT_h} r(t)\cos(2\pi f_j t)dt\right]^2 + \left[\int_{(i-1)T_h}^{iT_h} r(t)\sin(2\pi f_j t)dt\right]^2 \qquad (10\text{-}16)$$

Equations (10-8), (10-15), and (10-16) define the optimum noncoherent detector for a fast frequency-hopping signal with MFSK. Performance analyses for the detectors of fast frequency-hopping signals are given in [33].

Instead of basing detector design on the average likelihood ratio, one might apply a composite hypothesis test in which the presence of the signal is detected while simultaneously one or more of the unknown parameters under hypothesis H_1 are estimated. To simultaneously detect the signal and determine the frequency-hopping pattern, (10-8) is replaced by the *generalized likelihood ratio*:

$$\Lambda\left[r(t)\right] = \prod_{i=1}^{N_h} \max_{i\le j\le M}\left\{\Lambda_{ij}\left[r(t)|f_j\right]\right\} \qquad (10\text{-}17)$$

where the equations and subsystems for $\Lambda_{ij}[r(t)|f_j]$ remain the same. Equation (10-17) indicates that a maximum-likelihood estimate of f_j is made for each hop. Thus, an optimum test to determine the frequency channel occupied by the frequency-hopping signal is conducted during each hop. Although the detection performance is suboptimum when the generalized likelihood ratio is used to design a detector, this detector provides an important signal feature and is slightly easier to implement and analyze [32], [33].

10.2 Wideband Radiometer

Among the many alternatives to the optimum detector, two of the most useful are the *wideband radiometer* and the *channelized radiometer*. The wideband radiometer [24] is notable in that it requires virtually no detailed information about the parameters of the frequency-hopping signals to be detected other than their rough spectral location. The price paid for this robustness is much worse performance than more sophisticated detectors that exploit additional information about the signal [32]. The channelized radiometer is designed to explicitly exploit the spectral characteristics of frequency-hopping signals. In its optimal form, the channelized radiometer gives a performance nearly as good as that of the optimum detector. In its suboptimal form, the channelized radiometer trades performance for practicality and the easing of the required *a priori* information about the signal to be detected.

Suppose that the signal to be detected is approximated by a zero-mean, white Gaussian process. Consider two hypotheses that both assume the presence of a zero-mean, white Gaussian process over an observation interval $0 \leq t \leq T$. Under H_0 only noise is present, and the power spectral density is N_0, while under H_1 both signal and noise are present, and the power spectral density is N_1. Using N orthonormal basis functions as in the derivation of (10-4) and (10-5), we find that the conditional densities are

$$f(\mathbf{r}|H_i) = \prod_{l=1}^{N} \frac{1}{\sqrt{\pi N_i}} \exp\left(-\frac{r_l^2}{N_i}\right), \quad i = 0, 1 \tag{10-18}$$

Calculating the likelihood ratio, taking the logarithm, and merging constants with the threshold, we find that the decision rule is to compare

$$V = \sum_{l=1}^{N} r_l^2 \tag{10-19}$$

to a threshold. If we let $N \rightarrow \infty$ and use the properties of orthonormal basis functions, then we find that the test statistic is

$$V = \int_0^T r^2(t)dt \tag{10-20}$$

which defines an *energy detector* or *radiometer*. Although it was derived for a white Gaussian process, the radiometer is a reasonable configuration for determining the presence of unknown deterministic signals.

A radiometer may have one of the three equivalent forms shown in Figure 44. Consider the system of Figure 44a, which gives a direct realization of (10-20). The bandpass filter is assumed to be an ideal rectangular filter with center frequency f_c, bandwidth W, and output

$$r(t) = s(t) + n(t) \tag{10-21}$$

where $s(t)$ is a deterministic signal, and $n(t)$ is bandlimited white Gaussian noise with a two-sided power spectral density equal to $N_0/2$. Substituting (10-21) into (10-20), taking the expected value, and observing that $n(t)$ is a zero-mean process, we obtain

$$\begin{aligned} E[V] &= \int_0^T s^2(t)dt + \int_0^T E[n^2(t)]dt \\ &= \mathcal{E} + N_0 TW \end{aligned} \tag{10-22}$$

which indicates that the radiometer output is an unbiased estimate of the total energy in $r(t)$.

According to the results of Appendix B-2, a bandlimited deterministic signal can be represented as

$$s(t) = s_c(t)\cos 2\pi f_c t - s_s(t)\sin 2\pi f_c t \tag{10-23}$$

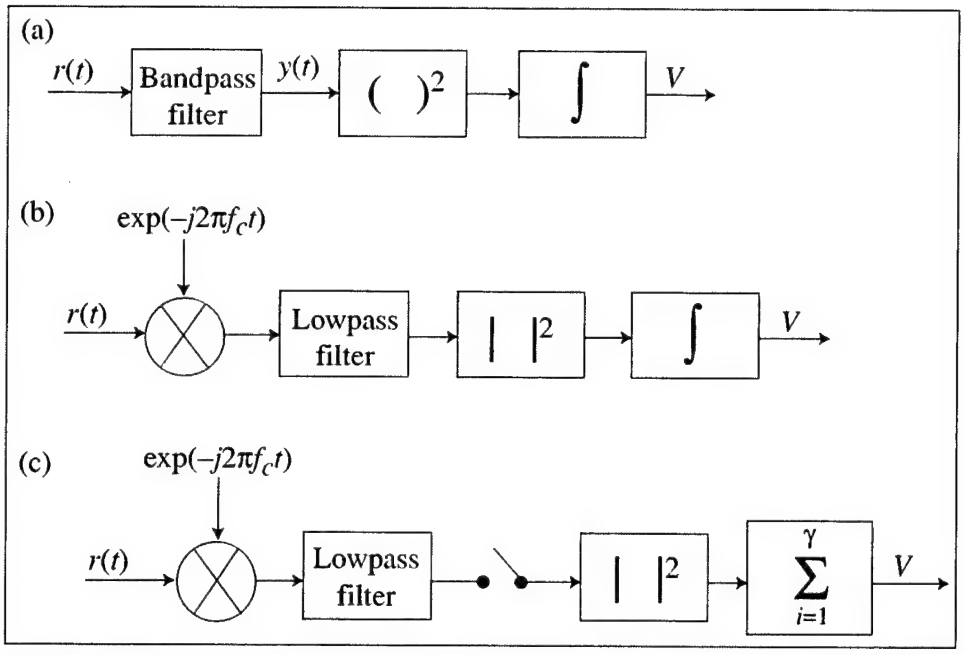


Figure 44. Radiometers: (a) passband, (b) baseband with integration, and (c) baseband with sampling and summation.

Since the spectrum of $s(t)$ is confined within the filter passband, $s_c(t)$ and $s_s(t)$ have frequency components confined to the band $|f| \leq W/2$. The Gaussian noise emerging from the bandpass filter can be represented in terms of quadrature components as (Appendix B-3)

$$n(t) = n_c(t) \cos 2\pi f_c t - n_s(t) \sin 2\pi f_c t \tag{10-24}$$

where $n_c(t)$ and $n_s(t)$ have flat power spectral densities, each equal to N_0 over $|f| \leq W/2$. Substituting (10-24), (10-23), and (10-21) into (10-20) and assuming that $f_c >> W$ and $f_c >> 1/T$, we obtain

$$V = \frac{1}{2}\int_0^T [s_c(t) + n_c(t)]^2\, dt + \frac{1}{2}\int_0^T [s_s(t) + n_s(t)]^2\, dt \tag{10-25}$$

A straightforward calculation verifies that the baseband radiometer of Figure 44(b) also produces this test statistic.

The *sampling theorems* for deterministic and stochastic processes (Appendix B-4) provide expansions of $s_c(t)$, $s_s(t)$, $n_c(t)$ and $n_s(t)$ that facilitate a statistical performance analysis. For example,

$$s_c(t) = \sum_{i=-\infty}^{\infty} s_c\left(\frac{i}{W}\right) \text{sinc}(\text{Wt} - \text{i}) \tag{10-26}$$

where sinc $x = (\sin \pi x)/\pi x$. Since the Fourier transform of the sinc function is a rectangular function, using Parseval's theorem from Fourier analysis and evaluating the resulting integral yields the approximations:

$$\int_0^T \text{sinc}^2(Wt-i)dt \approx \int_{-\infty}^{\infty} \text{sinc}^2(Wt-i)dt = \frac{1}{W}, \quad 0 < i \leq TW \tag{10-27}$$

$$\int_0^T \text{sinc}(Wt-i)\text{sinc}(Wt-j)dt \approx \int_{-\infty}^{\infty} \text{sinc}(Wt-i)\text{sinc}(Wt-j)dt = 0, \quad i \neq j \tag{10-28}$$

The rapid decline of sinc x for $|x| > 1$ implies that

$$\int_0^T \text{sinc}^2(Wt-i)dt \approx 0, \quad i \leq 0 \text{ or } i > TW \tag{10-29}$$

Substituting expansions similar to (10-26) into (10-25) and then using the preceding approximations, we obtain

$$V \approx \frac{1}{2W}\sum_{i=1}^{\gamma}\left[s_c\left(\frac{i}{W}\right)+n_c\left(\frac{i}{W}\right)\right]^2 + \frac{1}{2W}\sum_{i=1}^{\gamma}\left[s_s\left(\frac{i}{W}\right)+n_s\left(\frac{i}{W}\right)\right]^2 \tag{10-30}$$

where $\gamma = \lfloor TW \rfloor$ and it is always assumed that $TW \geq 1$. The error introduced by (10-29) at $i = 0$ and the error introduced by (10-27) at $i = TW$ are both nearly $1/2W$. For other values of i, the errors caused by the approximations are much less than $1/2W$ and decrease as TW increases. Equation (10-30) becomes an increasingly accurate approximation of (10-25) as γ increases. A test statistic proportional to (10-30) can be derived for the baseband radiometer of Figure 44(c) without invoking the sampling theorems and the accompanying approximations.

Because $n(t)$ is a zero-mean Gaussian process and has a power spectral density that is symmetrical about f_c, $n_c(t)$ and $n_s(t)$ are zero-mean, independent Gaussian processes (Appendix B-3). Thus, $n_c(i/W)$ and $n_s(j/W)$ are zero-mean, independent Gaussian random variables. Equation (B-40) implies that the power spectral densities of $n_c(t)$ and $n_s(t)$ are

$$S_c(f) = S_s(f) = \begin{cases} N_0, & |f| \leq W/2 \\ 0, & |f| > W/2 \end{cases} \tag{10-31}$$

The associated autocorrelation functions are

$$R_c(\tau) = R_s(\tau) = N_0 W \text{sinc}(W\tau) \tag{10-32}$$

This expression indicates that $n_c(i/W)$ is statistically independent of $n_c(j/W)$, $i \neq j$, and similarly for $n_s(i/W)$ and $n_s(j/W)$. Therefore, (10-30) becomes

$$V = \frac{N_0}{2}\left\{\sum_{i=1}^{\gamma} A_i^2 + \sum_{i=1}^{\gamma} B_i^2\right\} \tag{10-33}$$

where the A_i and the B_i are statistically independent Gaussian random variables with unit variances and means

$$m_{1i} = E[A_i] = \frac{1}{\sqrt{N_0 W}} s_c\left(\frac{i}{W}\right) \tag{10-34}$$

$$m_{2i} = E[B_i] = \frac{1}{\sqrt{N_0 W}} s_s\left(\frac{i}{W}\right) \tag{10-35}$$

Thus, $2V/N_0$ has a *noncentral chi-squared (χ^2) distribution* [24] with 2γ degrees of freedom and a noncentral parameter

$$\begin{aligned}\lambda &= \sum_{i=1}^{\gamma} m_{1i}^2 + \sum_{i=1}^{\gamma} m_{2i}^2 = \frac{1}{N_0 W}\sum_{i=1}^{\gamma} s_c^2\left(\frac{i}{W}\right) + \frac{1}{N_0 W}\sum_{i=1}^{\gamma} s_s^2\left(\frac{i}{W}\right) \\ &\approx \frac{1}{N_0}\int_0^T \left[s_c^2(t) + s_s^2(t)\right] dt \approx \frac{2}{N_0}\int_0^T s^2(t)dt = \frac{2\mathcal{E}}{N_0}\end{aligned} \tag{10-36}$$

The probability density function of $Z = 2V/N_0$ is

$$f_Z(x) = \frac{1}{2}\left(\frac{x}{\lambda}\right)^{(\gamma-1)/2} \exp\left(-\frac{x+\lambda}{2}\right) I_{\gamma-1}\left(\sqrt{x\lambda}\right) u(x) \tag{10-37}$$

where $I_n(\,)$ is the modified Bessel function of the first kind and order n, and $u(x)$ = 1, $x \geq 0$, and $u(x)$ = 0, $x < 0$. Using the series expansion in λ of the Bessel function and then setting λ = 0 in (10-37), we obtain the probability density function for Z in the absence of the signal:

$$f_Z(x) = \frac{1}{2^\gamma \Gamma(\gamma)} x^{\gamma-1} \exp\left(-\frac{x}{2}\right) u(x)\,, \quad \lambda = 0 \tag{10-38}$$

where $\Gamma(x)$ is the gamma function.

By straightforward calculations using the statistics of Gaussian variables, (10-33) and the subsequent results yield

$$E[V] = \mathcal{E} + N_0\gamma \tag{10-39}$$

$$\text{var}(V) = 2N_0\mathcal{E} + N_0^2\gamma \tag{10-40}$$

Equation (10-39) approaches the exact result of (10-22) as TW increases.

Let V_t denote the threshold level to which V is compared. A false alarm occurs if $V > V_t$ when the signal is absent. Application of (10-38) yields the probability of a false alarm:

$$\begin{aligned}P_F &= \int_{2V_t/N_0}^{\infty} \frac{1}{2^\gamma \Gamma(\gamma)} v^{\gamma-1} e^{-v/2} dv \\ &= 1 - \Gamma\left(\frac{2V_t}{N_0}, \gamma\right)\end{aligned} \tag{10-41}$$

where the *incomplete gamma function* is defined as

$$\Gamma(x,a) = \frac{1}{\Gamma(a)} \int_0^x e^{-t} t^{a-1} dt \tag{10-42}$$

and $\Gamma(\infty, a) = \Gamma(a)$. Integrating (10-41) by parts $\gamma - 1$ times yields the series

$$P_F = \exp\left(-\frac{V_t}{N_0}\right) \sum_{i=0}^{\gamma-1} \frac{1}{i!} \left(\frac{V_t}{N_0}\right)^i \tag{10-43}$$

Since correct detection occurs if $V > V_t$ when the signal is present, (10-37) indicates that the probability of detection is

$$P_D = \int_{2V_t/N_0}^{\infty} \frac{1}{2} \left(\frac{v}{\lambda}\right)^{(\gamma-1)/2} \exp\left(-\frac{v+\lambda}{2}\right) I_{\gamma-1}\left(\sqrt{v\lambda}\right) dv \tag{10-44}$$

The *generalized Marcum Q-function* is defined as

$$Q_m(\alpha,\beta) = \int_\beta^\infty x \left(\frac{x}{\alpha}\right)^{m-1} \exp\left(-\frac{x^2+\alpha^2}{2}\right) I_{m-1}(\alpha x) dx \tag{10-45}$$

where m is a nonnegative integer, and α and β are nonnegative real numbers. A change of variables (10-44) yields

$$P_D = Q_\gamma\left(\sqrt{\lambda}, \sqrt{2V_t/N_0}\right) \tag{10-46}$$

The threshold V_t is usually set to a value that ensures a specified P_F. To derive an easily computed closed-form expression for V_t in terms of P_F, we first approximate (10-41). When $TW >> 1$, $\gamma \approx TW$, and the central limit theorem for the sum of independent, identically distributed random variables with finite means and variances indicates that the distribution of V given by (10-33) is approximately Gaussian. Using (10-39) and (10-40) with $\mathcal{E} = 0$ and the Gaussian distribution, we obtain

$$\begin{aligned} P_F &\approx \frac{1}{(2\pi N_0^2 TW)^{1/2}} \int_{V_t}^{\infty} \exp\left[-\frac{(v - N_0 TW)^2}{2N_0^2 TW}\right] dv \\ &= Q\left[\frac{V_t - N_0 TW}{(N_0^2 TW)^{1/2}}\right], \quad TW >> 1 \end{aligned} \tag{10-47}$$

where $Q(x)$ is defined by (A-30). Inverting this equation, we obtain V_t in terms of P_F and N_0. Accordingly, if we estimate the value of N_0 to be $\hat{N}_0$ and P_F is specified, then the threshold should be

$$V_t \approx \hat{N}_0 \sqrt{TW} Q^{-1}(P_F) + \hat{N}_0 TW, \quad TW >> 1 \tag{10-48}$$

where $Q^{-1}(\,)$ denotes the inverse of the function $Q(\,)$. In the absence of a signal, (10-22) indicates that $N_0 = E[V]/TW$. Thus, N_0 can be estimated by averaging sampled radiometer outputs when it is known that no signal is present.

In some applications, one might wish to specify the *false alarm rate*, which is the expected number of false alarms per unit time, rather than P_F. If successive observation intervals do not overlap each other except possibly at end points, then the false alarms rate is

$$F = \frac{P_F}{T} \tag{10-49}$$

For $TW > 100$, the generalized Marcum Q-function is difficult to compute. Thus, we seek an approximation that is easier to compute and to invert. If V is approximated by a Gaussian random variable, then (10-39) and (10-40) imply that

$$P_D \approx Q\left[\frac{V_t - N_0 TW - \mathcal{E}}{(N_0^2 TW + 2N_0\mathcal{E})^{1/2}}\right], \quad TW >> 1 \tag{10-50}$$

Figure 45 depicts P_D versus $\mathcal{E}/N_0$ for a wideband radiometers with $\hat{N}_0 = N_0$ and $P_F = 10^{-3}$. Equations (10-48) and (10-50) are used to calculate V_t and P_D, respectively. The figure illustrates the increased energy required to maintain a specified P_D as TW increases. The figure also illustrates the impact of the imperfect estimation of N_0 when $P_F = 10^{-3}$ and $TW = 10^7$. When the estimation uncertainty is enough that $\hat{N}_0 = 1.001\ N_0$, the detection probability is lowered considerably.

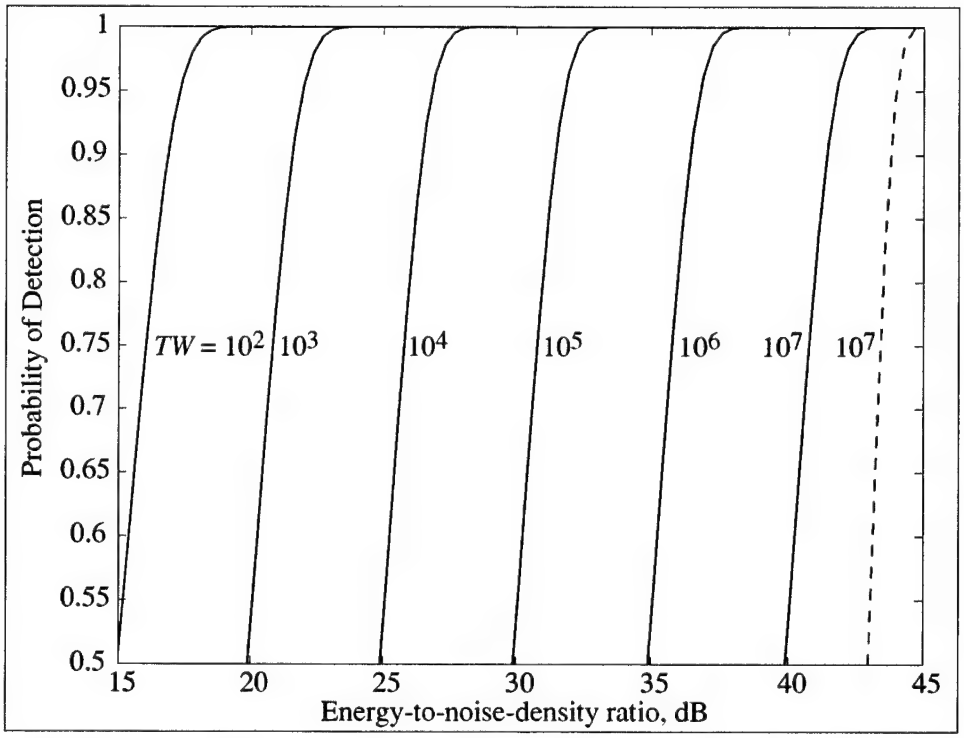


Figure 45. Probability of detection versus $\mathcal{E}/N_0$ for wideband radiometer with $P_F = 10^{-3}$ and various values of TW. Solid curves are the $\hat{N}_0 = N_0$; dashed curve is for $\hat{N}_0 = 1.001 N_0$.

The sensitivity of the radiometer to errors in $\hat{N}_0$ when TW is large, which has been observed experimentally [34], is due to the fact the $\mathcal{E}[V]$ contains a bias term equal to N_0TW and var(V) contains a term equal to $N_0^2\ TW$, as indicated by (10-39) and (10-40). Setting $\hat{N}_0$ high enough that $\hat{N}_0 \geq N_0$ ensures that V_t will be large enough that the required P_F is achieved regardless of the exact value of N_0. It is important that $\hat{N}_0/N_0$ is as close to unity as possible to avoid degrading P_D when TW is large. Consequently, the radiometer output due to noise alone, which provides $\hat{N}_0$, should be observed often enough that $\hat{N}_0$ closely tracks the changes in N_0 that might result from small changes in the circuitry or the environmental noise.

When V_t is specified, the value of $\mathcal{E}/N_0$ necessary to achieve a specified value of P_D may be obtained by inverting (10-44), which is computationally difficult but can be closely approximated [35]. The approximation (10-50) yields the necessary value

$$\frac{\mathcal{E}}{N_0} \approx \left[Q^{-1}(P_D)\right]^2 + \frac{V_t}{N_0} - TW + |Q^{-1}(P_D)|\sqrt{\left[Q^{-1}(P_D)\right]^2 + \frac{2V_t}{N_0} - TW}\ , \quad TW >> 1 \tag{10-51}$$

If V_t is approximated by (10-48), then a substitution into (10-51) and a rearrangement of terms yields

$$\frac{\mathcal{E}}{N_0} \approx h\sqrt{TW}\beta + (h-1)TW + \psi(\beta, \xi, TW, h), \quad TW >> 1 \tag{10-52}$$

where

$$\beta = Q^{-1}(P_F), \quad \xi = Q^{-1}(P_D), \quad h = \hat{N}_0/N_0 \tag{10-53}$$

$$\psi(\beta, \xi, TW, h) = \xi^2 + \sqrt{TW}|\xi|\left[2h - 1 + \frac{2\beta h}{\sqrt{TW}} + \frac{\xi^2}{TW}\right]^{1/2} \tag{10-54}$$

As TW increases, the significance of the third term in (10-52) decreases, while that of the second term increases if $h > 1$. Figure 46 shows $\mathcal{E}/N_0$ versus TW for $P_D = 0.99$ and various values of P_F and h.

Denoting the intercepted signal power by S and the signal duration by T_1, we find from (10-53) with $\mathcal{E} = ST_1$ that the intercepted power necessary to achieve specified values of P_D and either P_F or F is

$$\frac{S}{N_0} \approx \begin{cases} h\dfrac{\sqrt{TW}}{T_1}\beta + (h-1)W\dfrac{T}{T_1} + \dfrac{\psi}{T_1}\ , & T_1 < T \\ h\sqrt{\dfrac{W}{T}}\beta + (h-1)W + \dfrac{\psi}{T}\ , & T_1 \geq T \end{cases} \tag{10-55}$$

As long as $T_1 \geq T$, this equation indicates that increasing the observation interval decreases the required power. However, if $T_1 < T$, an increase in the

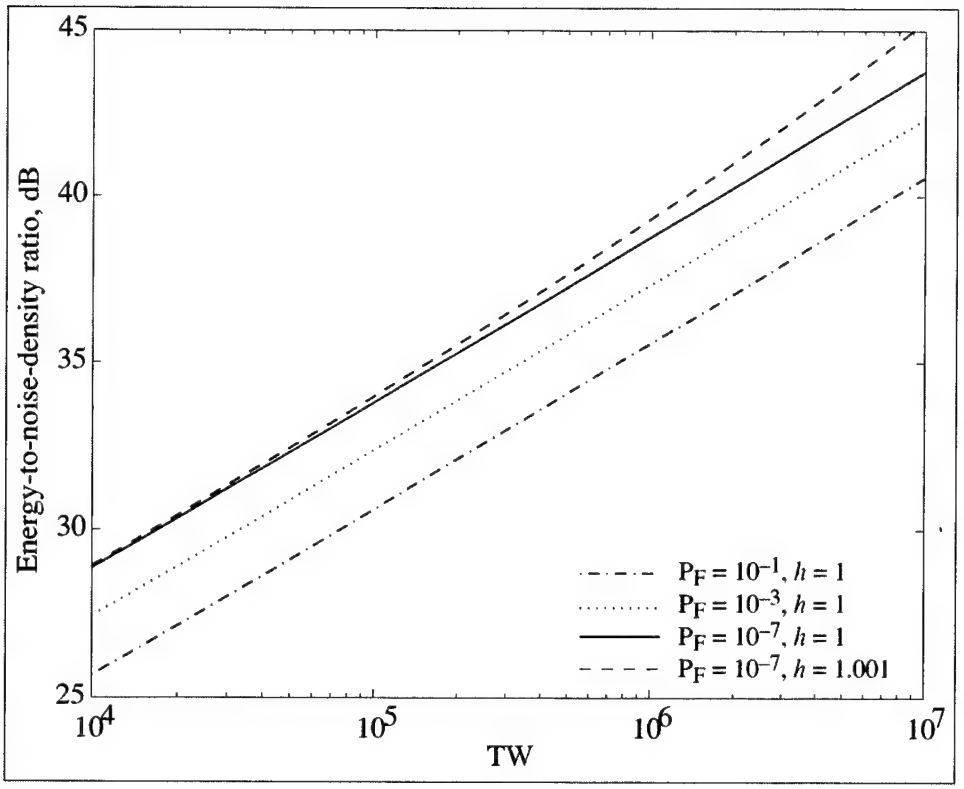


Figure 46. Energy-to-noise-density ratio versus TW for wideband radiometer with P_D = 0.99 and various values of P_F and h.

observation interval increases the required power. The required power determines the transmitter range that can be accommodated by the wideband radiometer. If the outputs of ν independent radiometers are averaged, the required power is reduced by the factor $1/\sqrt{\nu}$.

10.3 Channelized Radiometer

A channelized radiometer comprises K parallel radiometers, each of which has the form of Figure 44 and monitors a disjoint portion of the hopping band of a frequency-hopping signal, as depicted in Figure 47. The largest of the sampled radiometer outputs is compared to a threshold V_t stored in a comparator. If the threshold is exceeded, the comparator sends a 1 to the summer; otherwise it sends a 0. If the hop dwell epochs are at least approximately known, the channelized radiometer may improve its detection reliability by adding the 1's produced by N consecutive comparator outputs corresponding to multiple frequency hops of the signal to be detected. A signal is declared to be present if the sum V equals or exceeds the integer r, which serves as a second threshold. The two thresholds V_t are r are jointly optimized for the best system performance.

Ideally, $K = M$, the number of frequency channels in a hopset, but many fewer radiometers may be a practical or economic necessity; if so, each radiometer may monitor M_r frequency channels, where $1 \leq M_r \leq M$. Because of insertion

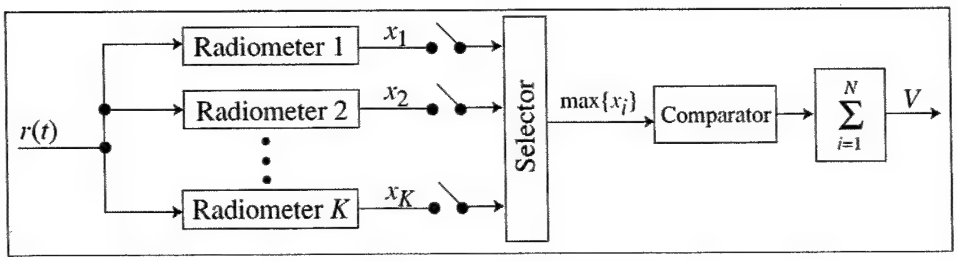


Figure 47. Channelized radiometer.

losses and the degradation caused by a power divider, it is unlikely that many more than 30 parallel radiometers are practical. An advantage of each radiometer covering many frequency channels is the reduced sensitivity to imprecise knowledge of the spectral boundaries of frequency channels. Since it is highly desirable to implement the parallel radiometers with similar circuitry, their bandwidths are assumed to be identical henceforth.

To prevent steady interference in a single radiometer from causing false alarms, the channelized radiometer must be able to recognize when one of its constituent radiometers produces an output above the threshold for too many consecutive samples. The channelized system may then delete that constituent radiometer's output from the detection algorithm or it may reassign the radiometer to another spectral location.

In the subsequent analysis of the channelized radiometer of Figure 46, the observation interval of the parallel radiometers, which is equal to the sampling interval, is assumed to equal the hop duration T_h. The effective observation time of the channelized radiometer, $T = NT_h$, should be less than the minimum expected message duration to avoid processing extraneous noise. Let P_{F1} denote the probability that a particular radiometer output at the sampling time exceeds the comparator threshold V_t when no signal is present. This probability is given by the right-hand side of (10-41). Therefore, (10-48) implies that if the sampling times are aligned with the frequency transitions, then the threshold necessary to achieve a specified P_{F1} is

$$V_t \approx \hat{N}_0 \sqrt{M_r T_h B} Q^{-1}(P_{F1}) + \hat{N}_0 M_r T_h B \ , \quad M_r T_h B >> 1 \tag{10-56}$$

where B is the bandwidth of each of the M_r frequency channels encompassed by a radiometer passband. The probability that at least one of the K parallel radiometer outputs exceeds V_t is

$$P_{F2} = 1 - (1 - P_{F1})^K \tag{10-57}$$

assuming that the channel noises are statistically independent because the radiometer passbands are disjoint. The probability of a false alarm of the channelized radiometer is the probability that the output V exceeds a threshold r:

$$P_F = \sum_{i=r}^{N} \binom{N}{i} P_{F2}^i (1 - P_{F2})^{N-i} \tag{10-58}$$

To solve this equation for P_{F2} in terms of P_F, we observe that the *incomplete beta function* is defined as

$$F(x|a,b) = \frac{1}{B(a,b)} \int_0^x t^{a-1}(1-t)^{b-1}dt\ , \quad 0 \le x \le 1 \tag{10-59}$$

where $B(a,b)$ is the *beta function* and $F(1|a,b) = 1$. In terms of this function, (10-58) becomes

$$P_F = F(P_{F2}|r, N-r+1) \tag{10-60}$$

The inverse of the incomplete beta function, which we denote by F^{-1}, may be easily computed by Newton's method or approximations [35]. Therefore, (10-56), (10-57), and (10-60) may be combined to determine the approximate threshold necessary to achieve a specified P_F:

$$V_t \approx \hat{N}_0\sqrt{M_rT_hB}Q^{-1}\left\{1-\left[1-F^{-1}\left(P_F|r,N-r+1\right)\right]^{1/K}\right\} + \hat{N}_0M_rT_hB\ , \quad M_rT_hB >> 1 \tag{10-61}$$

where it is assumed that N_0 does not vary across the hopping band, and hence there is one $\hat{N}_0$ and one V_t for all the parallel radiometers.

If the intercepted signal duration, T_1, is less than the observation time T, we assume for simplicity that $N_1 = T_1/T_h$, the number of sampling intervals during which the signal is present, is an integer. Furthermore, we assume that at most a single radiometer contains the intercepted signal during each sampling interval. Let P_{D1} denote the probability that a particular radiometer output exceeds the threshold when a signal is present in that radiometer. Thus, (10-46), (10-36) and (10-50) imply that

$$\begin{aligned} P_{D1} &= Q_L\left(\sqrt{\mathcal{E}_h/N_0}\ ,\ \sqrt{2V_t/N_0}\right) \\ &\approx Q\left[\frac{V_t - N_0M_rT_hB - \mathcal{E}_h}{\left(N_0^2M_rT_hB + 2N_0\mathcal{E}_h\right)^{1/2}}\right]\ , \quad M_rT_hB >> 1 \end{aligned} \tag{10-62}$$

where $L = \lfloor M_rT_hB \rfloor$ and $\mathcal{E}_h$ is the energy per hop dwell time. Let P_{D2} denote the probability that the threshold is exceeded by the sampled maximum of the parallel radiometer outputs. It is assumed that when a signal is present it occupies any one of M frequency channels with equal probability and that all radiometer passbands are within the hopping band. Consequently, the signal has probability M_r/M of being in the passband of a particular radiometer and probability KM_r/M of being in the passband of some radiometer. Since a detection may be declared in response to a radiometer that does not receive the signal,

$$P_{D2} = \frac{KM_r}{M}\left[1-(1-P_{D1})(1-P_{F1})^{K-1}\right] + \left(1-\frac{KM_r}{M}\right)P_{F2} \tag{10-63}$$

where the second term vanishes if the radiometer passbands cover the hopping band so that $KM_r = M$. The probability of detection associated with the observation interval when the signal is actually present during $N_1 \leq N$ of the hop intervals is

$$P_D = \sum_{i=r}^{N} \sum_{j=0}^{i} \binom{N_1}{j} \binom{N - N_1}{i - j} P_{D2}^{j} (1 - P_{D2})^{N_1 - j} P_{F2}^{i-j} (1 - P_{F2})^{N - N_1 - i + j} \tag{10-64}$$

If the minimum duration of a frequency-hopping signal is known, the overestimation of N might be avoided so that $N_1 = N$. The detection probability then becomes

$$\begin{aligned} P_D &= \sum_{i=r}^{N} \binom{N}{i} P_{D2}^{i} (1 - P_{D2})^{N-i} \\ &= F(P_{D2} | r, N - r + 1) \end{aligned} \tag{10-65}$$

A suitable, but not optimal, choice for the second threshold is $r = \lfloor N/2 \rfloor$ when the full hopping band is monitored by the channelized radiometer. In general, numerical results indicate that

$$r = \left\lfloor \frac{KM_r N}{2M} \right\rfloor \tag{10-66}$$

is a good choice for partial-band monitoring.

If detection decisions are made in terms of fixed observation intervals of duration $T = NT_h$, and successive intervals do not overlap except possibly at end points, then the false alarm rate defined in (10-49) is an appropriate design parameter. This type of detection is called *block detection*, and the false-alarm rate is

$$F = \frac{P_F}{NT_h} \tag{10-67}$$

To prevent the risk of major misalignment of the observation interval with the time the signal is being transmitted, either block detection must be supplemented with hardware for arrival-time estimation or the duration of successive observation intervals should be less than roughly half the anticipated signal duration.

A different approach to mitigating the effect of a misalignment, called *binary moving-window detection*, is for the observation interval to be constructed by dropping the first sampling interval of the preceding observation interval and adding a new sampling interval. A false alarm is considered to be a detection declaration at the end of the new interval when no signal is actually present. Thus, a false alarm occurs only if the comparator input for an added sampling interval exceeds the threshold, the comparator input for the discarded sampling

interval did not, and the count for the preceding observation interval was $r-1$. Therefore, the probability of a false alarm is

$$P_{F0} = C(0,1)C(r-1, N-1)C(1,1) \tag{10-68}$$

where

$$C(i,N) = \binom{N}{i} P_{F2}^{i} (1-P_{F2})^{N-i} \quad , \quad i \leq N \tag{10-69}$$

It follows that the false-alarm rate is

$$F = \frac{P_{F0}}{T_h} = \frac{r}{NT_h} \binom{N}{r} P_{F2}^{r} (1-P_{F2})^{N+1-r} \tag{10-70}$$

Since the right-hand side of (10-69) is proportional to the first term of the series in (10-58)

$$F \leq \frac{rP_F}{NT_h} \tag{10-71}$$

A comparison of this inequality with (10-67) indicates that the false alarm rate less than r times as large for moving-window detection as it is for block detection. Thus, moving-window detection requires a slightly higher comparator threshold for the same false-alarm rate and, hence, slightly more signal power to detect a frequency-hopping signal. However, moving-window detection with $N \approx N_1 >> 1$ inherently limits the misalignment between the occurrence of the intercepted signal and the observation interval. For some observation interval, the misalignment is not more than $T_h/2$.

As an example, it is assumed that there are $M = 2400$ frequency channels, the signal duration is known and there is no misalignment so that $N_1 = N$, block detection is used so that (10-67) is applicable, $F = 10^{-7}/T_h$, $B = 250/T_h$, and $\hat{N}_0 = N_0$. Figure 48 plots P_D versus $\mathcal{E}_h/N_0$ for the channelized radiometer with full hopping-band coverage so that $M_r = M/K$, and several values of K and N. The figure also shows the results for a wideband radiometer with $TW = BT_hM$ $N = 6 \cdot 10^5 \cdot N$, and $N = 150$ or 750. It is observed that the channelized radiometer with $K = 30$ is much better than the wideband radiometer when $N = 150$, but loses its advantage for $P_D \leq 0.995$ when $N = 750$. The substantial advantage of the channelized radiometer with $K = M$ and $M_r = 1$ is apparent. As N increases, the channelized radiometer can retain its advantage over the wideband radiometer by increasing K accordingly.

In Figure 49, $N = N_1 = 150$ and $K = 30$, but M_r and $\hat{N}_0/N_0$ are variable. The fraction of the hopping band monitored by the channelized radiometer is denoted by $\beta = KM_r/M$. It is observed that when $\hat{N}_0 = 1.001\ N_0$, there is only a small performance loss for the channelized radiometer despite the fact that $TW = 9 \cdot 10^7$. The relative insensitivity of the channelized radiometer to small errors in $\hat{N}_0$ is a major advantage over the wideband radiometer. The figure

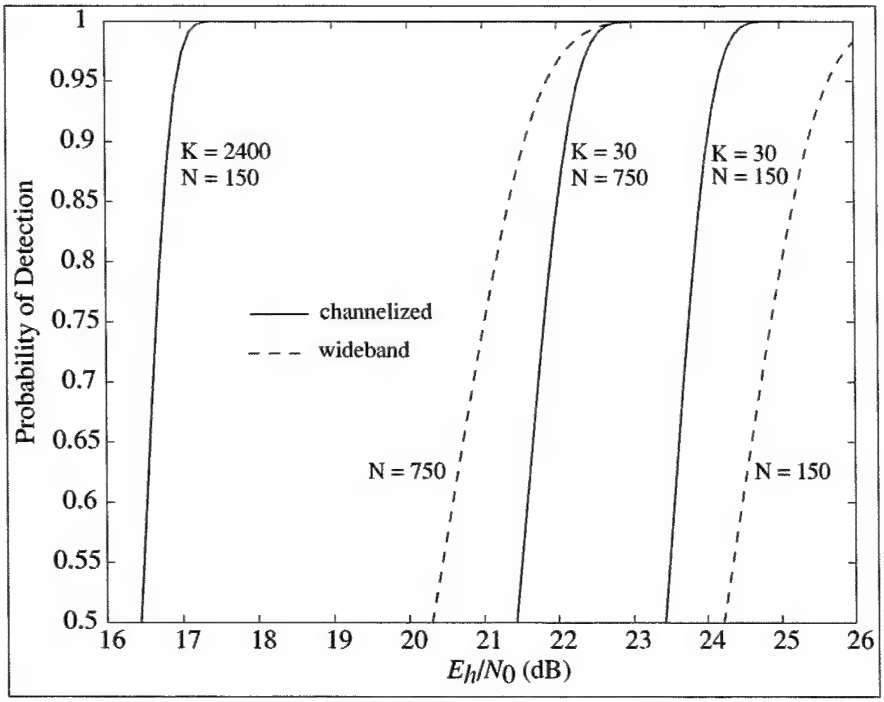


Figure 48. Probability of detection versus $\mathcal{E}_h/N_0$ for channelized and wideband radiometers with full coverage, $N_1 = N$, $\hat{N}_0 = N_0$, $M = 2400$, $F = 10^{-7}/T_h$, and $B = 250/T_h$.

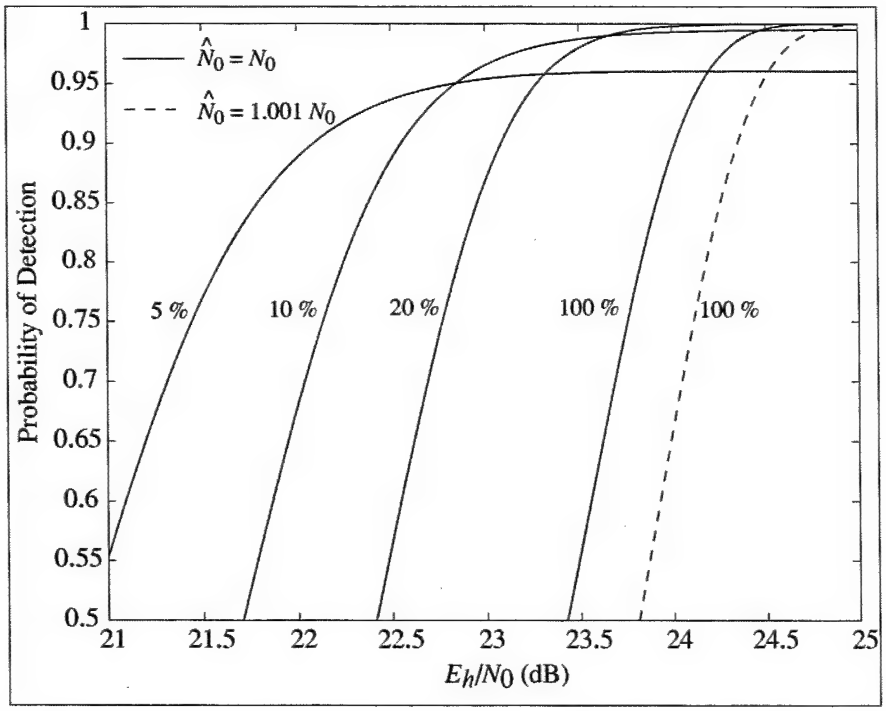


Figure 49. Probability of detection for channelized radiometer with different percentages of coverage, $N_1 = N = 150$, $K = 30$, $M = 2400$, $F = 10^{-7}/T_h$, $B = 250/T_h$, and $\hat{N}_0 = N_0$ or $\hat{N}_0 = 1.001\ N_0$.

illustrates the following tradeoff when K and M are fixed: as M_r decreases, fewer frequency channels are monitored, but less noise enters a radiometer. The net result is beneficial when β is reduced to 20 percent. However, the figure indicates that for β = 10 percent or 5 percent, the hopping-band coverage becomes inadequate to enable a P_D greater than 0.995 and 0.96, respectively, regardless of $\mathcal{E}_h/N_0$. Thus, there is a minimum fraction of the hopping band that must be monitored to ensure a specified P_D. Many other aspects of the channelized radiometer, including the effects of timing misalignments, are examined in [35].

Appendix A. Error-Correcting Codes

Error-correcting codes [36], [37], [38], [39] are vital in fully exploiting the potential capabilities of spread-spectrum communication systems. In this section, some of the fundamental results of coding theory are presented and used to derive the appropriate receiver computations and the error probabilities at receiver outputs.

A-1 Block Codes

An (n, k) *block code* is a set of *codewords*, each of which has n symbols and represents k information symbols. Each symbol is selected from an alphabet of q symbols, and there are q^k codewords. If $q = 2^m$, then an (n, k) code of q-ary symbols is equivalent to an (mn, mk) binary code. A block encoder can be implemented by using logic elements or memory to map a k-symbol word into an n-symbol codeword. After the waveform representing a codeword is received and demodulated, the decoder uses the demodulator output to determine the information symbols corresponding to the codeword. If the demodulator produces a sequence of discrete symbols and the decoding is based on these symbols, the demodulator is said to make *hard decisions*. Conversely, if the demodulator produces analog or multilevel quantized samples of the waveform, the demodulator is said to make *soft decisions*. The advantage of soft decisions is that reliability or quality information is provided to the decoder, which can use this information to improve its performance.

The *Hamming distance* between two sequences with an equal number of symbols is defined as the number of positions in which the symbol of one sequence differs from the corresponding symbol of the other sequence. The minimum Hamming distance between any two codewords is called the *minimum distance* of the code. When hard decisions are made, the demodulator output sequence is called the *received sequence* or the *received word*. Hard decisions imply that the overall channel between the encoder output and the decoder input is the classical binary symmetric channel. If the channel symbol error probability is less than one-half, then the maximum-likelihood criterion implies that the Hamming distance should be used to decide which codeword was transmitted. A *complete decoder*, which implements the maximum-likelihood criterion, assumes that the correct codeword is the one that is the smallest Hamming distance from the received word. An *incomplete decoder* does not attempt to correct all received words.

A conceptual three-dimensional representation of the vector space of sequences of length n is depicted in Figure A-1. Each codeword occupies the center of a *decoding sphere* with radius t in Hamming distance, where t is a positive integer.

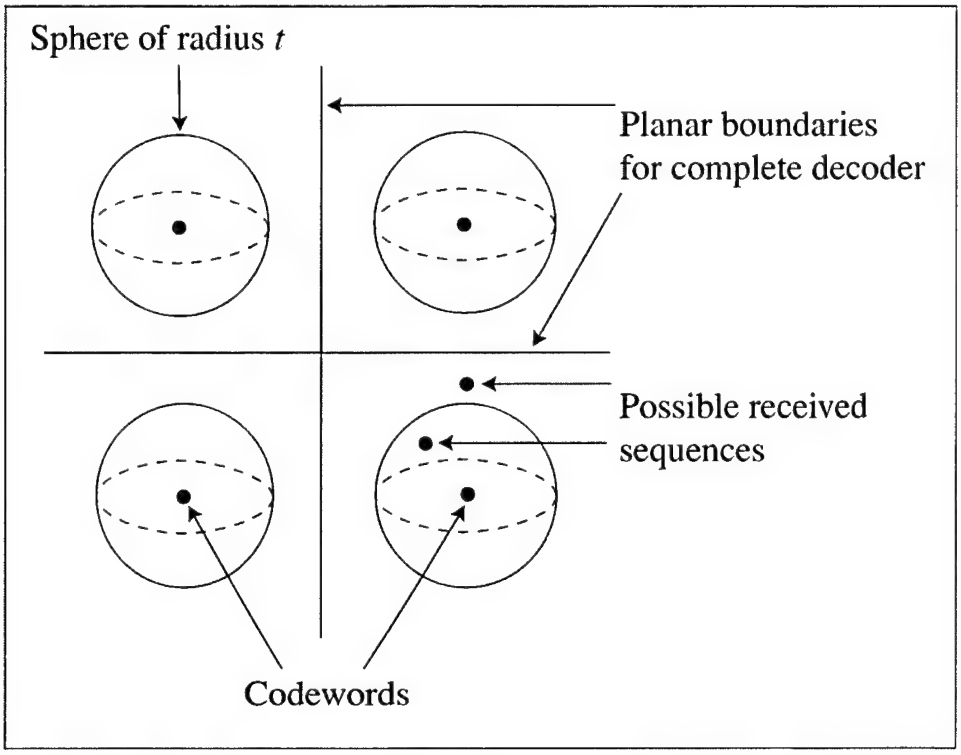


Figure A-1. Conceptual representation of vector space of sequences of length n.

For a complete decoder, planar boundaries define decision regions surrounding each codeword. A received word is assumed to be a corrupted version of the codeword enclosed by the boundaries. A *bounded-distance decoder* is an incomplete decoder that attempts to correct symbol errors in a received word if it lies within one of the decoding spheres. Since unambiguous decoding requires that none of the spheres may intersect, the maximum number of random errors that can be corrected by a bounded-distance decoder is

$$t = \lfloor (d_m - 1)/2 \rfloor \tag{A-1}$$

where d_m is the minimum Hamming distance between codewords and $\lfloor x \rfloor$ denotes the largest integer less than or equal to x. When more than t errors occur, the received word may lie within a decoding sphere surrounding an incorrect codeword or it may lie in the interstices (regions) outside the decoding spheres. If the received word lies within a decoding sphere, the decoder selects the incorrect codeword at the center of the sphere and produces an output word of information symbols with undetected errors. If the received word lies in the interstices, the decoder cannot correct the errors, but recognizes their existence. Thus, the decoder fails to decode the received word.

Since there are $\binom{n}{i}(q-1)^i$ words at exactly distance i from the center of the sphere, the number of words in a decoding sphere of radius t is determined from

elementary combinatorics to be

$$V = \sum_{i=0}^{t} \binom{n}{i} (q-1)^i \tag{A-2}$$

Since a block code has q^k codewords, $q^k V$ words are enclosed in some sphere. The number of possible received words is $q^n \geq q^k V$, which yields

$$q^{n-k} \geq \sum_{i=0}^{t} \binom{n}{i} (q-1)^i \tag{A-3}$$

This inequality with t given by (A-1) establishes the *Hamming upper bound* on the minimum distance.

A block code is called a *linear block code* if its codewords form a k-dimensional subspace of the vector space of sequences with n symbols. Thus, the vector sum or difference of two codewords of a linear block code is a codeword. If a binary block code is linear, the symbols of a codeword are modulo-two sums of information bits. Since a linear block code is a subspace of a vector space, it must contain the additive identity. Thus, the all-zero sequence is always a codeword in any linear block code. Since nearly all practical block codes are linear, block codes are assumed to be linear henceforth.

The *Hamming weight* of a codeword is defined as the number of nonzero symbols in a codeword. For binary block codes, the Hamming weight is the number of 1's in a codeword. For a linear block code, the sum of two codewords is another codeword with weight equal to the distance between the two original codewords. By adding the codeword $\mathbf{c}$ to all the codewords, we find that the set of Hamming distances from any codeword $\mathbf{c}$ is the same as the set of codeword weights. Consequently, in evaluating decoding error probabilities, one can assume without loss of generality that the all-zero codeword was transmitted, and the minimum Hamming distance is equal to the minimum weight of the nonzero codewords.

A *systematic block code* is a code in which the information symbols appear unchanged in the codeword, which also has additional parity symbols. In terms of the word error probability for the binary symmetric channel, every linear code is equivalent to a systematic linear code [36]. Therefore, systematic block codes are the standard choice and are assumed henceforth. Some systematic codewords have only one nonzero information symbol and $n-k$ parity symbols. These codewords have weights that cannot exceed $n-k+1$. Since the minimum distance of the code is equal to the minimum codeword weight,

$$d_m \leq n-k+1 \tag{A-4}$$

This upper bound, which is called the *Singleton bound*, is valid for all linear block codes. A code that has $d_m = n-k+1$ is called a *maximum-distance-separable code*.

A *cyclic code* is a linear block code in which a cyclic shift of the symbols of a codeword produces another codeword. This characteristic allows the implementation of encoders and decoders that use linear feedback shift registers. Relatively simple encoding and hard-decision decoding techniques are known for cyclic codes belonging to the class of *Bose-Chaudhuri-Hocquenghem* (BCH) *codes*, which may be binary or nonbinary. A BCH code has a length that is a divisor of $q^m - 1$, where $m \geq 2$, and is designed to have an error-correction capability of $t = \lfloor (\delta - 1)/2 \rfloor$, where δ is the *design distance*. Although the minimum distance may exceed the design distance, the standard BCH decoding algorithms cannot correct more than t errors. The parameters (n, k, t) for binary BCH codes with $7 \leq n \leq 127$ are listed in Table A-1.

Table A-1. Binary BCH codes.

n	k	t	D_p	n	k	t	D_p	n	k	t	D_p
7	4	1	1	63	45	3	0.1592	127	92	5	0.0077
7	1	3	1	63	39	4	0.0380	127	85	6	0.0012
15	11	1	1	63	36	5	0.0571	127	78	7	$1.68 \cdot 10^{-4}$
15	7	2	0.4727	63	30	6	0.0088	127	71	9	$2.66 \cdot 10^{-4}$
15	5	3	0.5625	63	24	7	0.0011	127	64	10	$2.48 \cdot 10^{-5}$
15	1	7	1	63	18	10	0.0044	127	57	11	$2.08 \cdot 10^{-6}$
31	26	1	1	63	16	11	0.0055	127	50	13	$1.42 \cdot 10^{-6}$
31	21	2	0.4854	63	10	13	0.0015	127	43	14	$9.11 \cdot 10^{-8}$
31	16	3	0.1523	63	7	15	0.0024	127	36	15	$5.42 \cdot 10^{-9}$
31	11	5	0.1968	63	1	31	1	127	29	21	$2.01 \cdot 10^{-6}$
31	6	7	0.1065	127	120	1	1	127	22	23	$3.56 \cdot 10^{-7}$
31	1	15	1	127	113	2	0.4962	127	15	27	$7.75 \cdot 10^{-7}$
63	57	1	1	127	106	3	0.1628	127	8	31	$8.10 \cdot 10^{-7}$
63	51	2	0.4924	127	99	4	0.0398	127	1	63	1

A *perfect code* is a block code such that every n-symbol sequence is at a distance of at most t from some n-symbol codeword, and the sets of all sequences at distance t or less from each codeword are disjoint. Thus, the Hamming bound of (A-3) is satisfied with equality, and a complete decoder is also a bounded-distance decoder. The only perfect codes are the binary repetition codes of odd length, the Hamming codes, the binary Golay (23,12) code, and the ternary Golay (11,6) code.

Repetition codes have only one information bit represented by n binary code symbols. When n is odd, the $(n, 1)$ repetition codes are perfect codes with $d_m = n$ and $t = (n - 1)/2$. A hard-decision decoder makes a decision based on the state of the majority of the demodulated symbols. Although repetition codes are not efficient for the additive white Gaussian noise (AWGN) channel, they can improve the system performance for fading channels if the number of repetitions is properly chosen.

A *Hamming* (n, k) *code* is a BCH code that has $d_m = 3$, $t = 1$, and

$$n = \frac{q^{n-k} - 1}{q - 1} \tag{A-5}$$

A Hamming code is a perfect code that is capable of correcting all single errors. Binary Hamming codes are found in Table A-1. The *Golay (23,12) code* is a binary cyclic code that is a perfect code with $d_m = 7$ and $t = 3$.

The *extended Golay* (24,12) *code* is formed by adding an overall parity symbol to the Golay (23,12) code, thereby increasing the minimum distance to $d_m = 8$. As a result, some received sequences with four errors can be corrected with a complete decoder. The (24,12) code is often preferable to the (23,12) code because the *code rate*, which is defined as the ratio k/n, is exactly one-half, which simplifies the system timing.

In general, any (n, k) linear block code with an odd value of d_m can be converted into an $(n + 1, k)$ *extended code* by adding a parity symbol. The advantage of the extended code stems from the fact that the minimum distance of the extended code is increased by one, which improves the performance. However, the decoding complexity and code rate are usually changed insignificantly.

Nonbinary block codes can accommodate high data rates efficiently because decoding operations are performed at the symbol rate rather than the higher information-bit rate. *Reed-Solomon codes* are nonbinary BCH codes with $n = q - 1$ and are maximum-distance-separable codes with $d_m = n - k + 1$. For convenience in implementation, q is usually chosen so that $q = 2^m$, where m is the number of bits per symbol. Thus, $n = 2^m - 1$ and the code provides correction of 2^m-ary symbols. Most Reed-Solomon decoders are bounded-distance decoders with $t = \lfloor (d_m - 1)/2 \rfloor$.

The *weight distribution* of a code is a list of the number of codewords with each possible weight. The weight distributions of the Golay codes are list in Table A-2. Analytical expressions that can be used to compute the weight distribution are known in some cases. Let A_l denote the number of codewords with weight l. For a binary Hamming code, each A_l can be determined from the weight-enumerator polynomial

$$A(x) = \sum_{l=0}^{n} A_l x^l = \frac{1}{n+1}[(1+x)^n + n(1+x)^{(n-1)/2}(1-x)^{(n+1)/2}] \tag{A-6}$$

For a maximum-distance-separable code, $A_0 = 1$ and [37]

$$A_l = \binom{n}{l}(q-1)\sum_{i=0}^{l-d_m}(-1)^i\binom{l-1}{i}q^{l-i-d_m}, \quad d_m \le l \le n \tag{A-7}$$

The weight distribution of other codes can be determined by examining all valid codewords if the number of codewords is not too large for a computation.

Table A-2. Weight Distribution of Golay Codes.

Weight	Number of Codewords (23,12)	(24,12)
0	1	1
7	253	0
8	506	759
11	1288	0
12	1288	2576
15	506	0
16	253	759
23	1	0
24	0	1

Error probabilities for hard-decision decoding

There are two types of bounded-distance decoders: erasing decoders and reproducing decoders. They differ only in their actions following the detection of uncorrectable errors in a received word. An *erasing decoder* discards the received word and may initiate an automatic retransmission request. For a systematic block code, a *reproducing decoder* reproduces the information symbols of the received word as its output.

Let P_s denote the *channel-symbol error probability*, which is the probability of error in a demodulated code symbol. It is assumed that the channel-symbol errors are statistically independent and identically distributed, which is usually an accurate model for systems with appropriate symbol interleaving (Section A-3). Let P_w denote the *word error probability*, which is the probability that a received word is not decoded correctly due to both undetected errors and decoding failures. There are $\binom{n}{i}$ distinct ways in which i errors may occur among n symbols. Since a received sequence may have more than t errors but no information-symbol errors,

$$P_w \leq \sum_{i=t+1}^{n} \binom{n}{i} P_s^i (1-P_s)^{n-i} \tag{A-8}$$

for a reproducing decoder that corrects t or few errors. For an erasing decoder, (A-8) becomes an equality. For error-correcting decoders, t is given by (A-1) because it is pointless to make the decoding spheres smaller than the maximum allowed by the code. However, if a block code is used for both error correction and error detection, an erasing decoder is often designed with t less than the maximum. If a block code is used exclusively for error detection, then $t = 0$.

Conceptually, a complete decoder correctly decodes when the number of symbol errors exceeds t if the received sequence lies within the planar boundaries associated with the correct codeword, as depicted in Figure A-1. When a received sequence is equidistant from two or more codewords, a complete

decoder selects one of them according to some arbitrary rule. Thus, the word error probability for a complete decoder satisfies (A-8). If $P_s \leq 1/2$, a complete decoder is a maximum-likelihood decoder.

Let P_{ud} denote the probability of an undetected error, and let P_{df} denote the probability of a decoding failure. For a bounded-distance decoder

$$P_{ud} + P_{df} = \sum_{i=t+1}^{n} \binom{n}{i} P_s^i (1 - P_s)^{n-i} \tag{A-9}$$

Thus, it is easy to calculate P_{df} once P_{ud} is determined. Since the set of Hamming distances from a given codeword to the other codewords is the same for all given codewords of a linear block code, it is legitimate to assume for convenience in evaluating P_{ud} that the all-zero codeword was transmitted. If channel-symbol errors are statistically independent and occur with the same probability P_s, then the probability of an error in a specific set of i positions of the received word is

$$P_e(i) = P_s^i (1 - P_s)^{n-i} \tag{A-10}$$

For an undetected error to occur at the output of a bounded-distance decoder, the number of erroneous symbols must exceed t and the received word must lie within an incorrect decoding sphere. Let $N(l, i)$ is the number of sequences of Hamming weight i that lie within a decoding sphere of radius t associated with a particular codeword of weight l. Then $N(l, i)/(q-1)^i$ is the number of sets of these sequences such that in each set the nonzero symbols occur in i particular positions. Using (A-10), we obtain

$$\begin{aligned} P_{ud} &= \sum_{i=t+1}^{n} P_e(i) \sum_{l=\max(i-t,d_m)}^{\min(i+t,n)} A_l \frac{N(l,i)}{(q-1)^i} \\ &= \sum_{i=t+1}^{n} \left(\frac{P_s}{q-1}\right)^i (1 - P_s)^{n-i} \sum_{l=\max(i-t,d_m)}^{\min(i+t,n)} A_l N(l,i) \end{aligned} \tag{A-11}$$

Consider sequences of weight i that are at distance s from a particular codeword of weight l, where $|l - i| \leq s \leq t$ so that the sequences are within the decoding sphere of the codeword. By counting these sequences and then summing over the allowed values of s, we can determine $N(l, i)$. The counting is done by considering changes in the components of this codeword that can produce one of these sequences. Let j denote the number of nonzero codeword symbols that are changed to zeros, α the number of codeword zeros that are changed to any of the $(q-1)$ nonzero symbols in the alphabet, and β the number of nonzero codeword symbols that are changed to any of the other $(q-2)$ nonzero symbols. For a sequence at distance s to result, it is necessary that $0 \leq j \leq s$. The number of sequences that can be obtained by changing any j of the l nonzero symbols to zeros is $\binom{l}{j}$, where $\binom{b}{a} = 0$ if $a > b$. For a specified value of j, it is necessary that $\alpha = j + i - l$ to ensure a sequence of weight i. The number of sequences that

result from changing any α of the $n-l$ zeros to nonzero symbols is $\binom{n-l}{\alpha}$ $(q-1)^{\alpha}$. For a specified value of j and hence α, it is necessary that $\beta = s-j-\alpha = s+l-i-2j$ to ensure a sequence at distance s. The number of sequences that result from changing β of the $l-j$ remaining nonzero components is $\binom{l-j}{\beta}$ $(q-2)^{\beta}$, where $0^x = 0$ if $x \neq 0$ and $0^0 = 1$. Summing over the allowed values of s and j, we obtain

$$N(l,i) = \sum_{s=|l-i|}^{t} \sum_{j=0}^{s} \binom{l}{j} \binom{n-l}{j+i-l} \binom{l-j}{s+l-i-2j} \times (q-1)^{j+i-l}(q-2)^{s+l-i-2j} \tag{A-12}$$

Equations (A-11) and (A-12) allow the exact calculation of P_{ud}.

When $q = 2$, the only term in the inner summation of (A-12) that is nonzero has the index $j = (s+l-i)/2$ provided that this index is an integer and $0 \leq (s+l-i)/2 \leq s$. Using this result, we find that for binary codes,

$$N(l,i) = \sum_{s=|l-i|}^{t} \binom{n-l}{\frac{s+i-l}{2}} \binom{l}{\frac{s+l-i}{2}}, \quad q = 2 \tag{A-13}$$

The word error probability is a performance measure that is important primarily in applications for which only a decoded word completely without symbol errors is acceptable. When the utility of a decoded word degrades in proportion to the number of information bits that are in error, the *information-bit error probability* is frequently used as a performance measure. To evaluate it for block codes that may be nonbinary, we first examine the information-symbol error probability.

Let $P_{is}(j)$ denote the probability of an error in information symbol j at the decoder output. In general, it cannot be assumed that $P_{is}(j)$ is independent of j. The *information-symbol error probability*, which is defined as the unconditional error probability without regard to the symbol position, is

$$P_{is} = \frac{1}{k} \sum_{j=1}^{k} P_{is}(j) \tag{A-14}$$

The random variables Z_j, $j = 1, 2, \ldots, k$, are defined so that $Z_j = 1$ if information symbol j is in error and $Z_j = 0$ if it is correct. The expected number of information-symbol errors is

$$E[I] = E\left[\sum_{j=1}^{k} Z_j\right] = \sum_{j=1}^{k} E[Z_j] = \sum_{j=1}^{k} P_{is}(j) \tag{A-15}$$

where $E[\,]$ denotes the expected value. The *information-symbol error rate* is defined as $E[I]/k$. Equations (A-14) and (A-15) imply that

$$P_{is} = \frac{E[I]}{k} \tag{A-16}$$

Thus, the information-symbol error probability is equal to the information-symbol error rate.

Let $P_{ds}(j)$ denote the probability of an error in symbol j of the codeword chosen by the decoder or symbol j of the received sequence if a decoding failure occurs. The decoded-symbol error probability is

$$P_{ds} = \frac{1}{n} \sum_{j=1}^{n} P_{ds}(j) \tag{A-17}$$

A derivation similar to that of (A-16) yields

$$P_{ds} = \frac{E[D]}{n} \tag{A-18}$$

where $E[D]$ is the expected number of decoded-symbol errors. Thus, the decoded-symbol error probability is equal to the decoded-symbol error rate.

It can be shown [40] that for cyclic codes, the error rate among the information symbols in the output of a bounded-distance decoder is equal to the error rate among all the decoded symbols; that is,

$$P_{is} = P_{ds} \tag{A-19}$$

When this equation is valid, the calculation of P_{is} is significantly simplified because P_{ds} can be expressed in terms of the code weight distribution, whereas an exact calculation of P_{is} requires additional information.

An erasing decoder makes an error only if it fails to detect one. Therefore, $P_{ds} = P_{ud}$ and (A-11) implies that for an erasing decoder

$$P_{ds} = \sum_{i=t+1}^{n} \left(\frac{P_s}{q-1}\right)^i (1-P_s)^{n-1} \sum_{l=\max(i-t,d_m)}^{\min(i+t,n)} A_l N(l,i) \quad \text{(erasing decoder)} \tag{A-20}$$

The number of sequences of weight i that lie in the interstices outside the decoding spheres is

$$L(i) = (q-1)^i \binom{n}{i} - \sum_{l=\max(i-t,d_m)}^{\min(i+t,n)} A_l N(l,i)\,, \quad i \geq t+1 \tag{A-21}$$

where the first term is the total number of sequences of weight i, and the second term is the number of sequences of weight i that lie within incorrect decoding spheres. When i symbol errors in the received word cause a decoding failure, the decoded symbols in the output of a reproducing decoder contain i errors. Therefore, the decoded-symbol error rate is

$$P_{ds} = \sum_{t+1}^{n} P_e(i) \left[\sum_{l=\max(i-t,d_m)}^{\min(i+t,n)} A_l N(l,i) \frac{l}{n} + L(i) \frac{i}{n} \right] \quad \text{(reproducing decoder)} \tag{A-22}$$

Even if $P_{is} = P_{ds}$, two major problems still arise in calculating P_{is} from (A-20) or (A-22). The computational complexity may be prohibitive when n and q are large, and the weight distribution is unknown for many linear or cyclic block codes, Hence, approximations are highly desirable.

The *packing density* is defined as the ratio of the number of words in the q^k decoding spheres to the total number of sequences of length n. From (A-2), it follows that the packing density is

$$D_p = \frac{q^k}{q^n} \sum_{i=0}^{t} \binom{n}{i} (q-1)^i \tag{A-23}$$

For perfect codes, $D_p = 1$. If $D_p > 0.5$, undetected errors tend to occur more often then decoding failures, and the code is considered tightly packed. If $D_p < 0.1$, decoding failures predominate, and the code is considered loosely packed. The packing densities of binary BCH codes are listed in Table 1. The codes are tightly packed if $n = 7$ or 15. For $k > 1$ and $n = 31$, 63, or 127, the codes are tightly packed only if $t = 1$ or 2.

To approximate P_{is} for tightly packed codes, let $A(i)$ denote the event that i errors occur in a received sequence of n symbols at the decoder input. If the symbol errors are independent, the probability of this event is

$$P[A(i)] = \binom{n}{i} P_s^i (1-P_s)^{n-i} \tag{A-24}$$

Given event $A(i)$ for i such that $d_m \leq i \leq n$, it is plausible to assume that a reproducing bounded-distance decoder usually chooses a codeword with approximately i symbol errors. For i such that $t+1 \leq i \leq d_m$, it is plausible to assume that the decoder usually selects a codeword at the minimum distance d_m. These approximations, (A-19), (A-24), and the identity $\binom{n}{i}\frac{i}{n} = \binom{n-1}{i-1}$ indicate that P_{is} for reproducing decoders is approximated by

$$P_{is} \approx \sum_{i=t+1}^{d_m} \frac{d_m}{n} \binom{n}{i} P_s^i (1-P_s)^{n-i} + \sum_{i=d_m+1}^{n} \binom{n-1}{i-1} P_s^i (1-P_s)^{n-i} \tag{A-25}$$

The virtues of this approximation are its lack of dependence on the code weight distribution and its generality. Computations for specific codes indicate that the accuracy of this approximation tends to increase with P_{ud}/P_{df}. The right-hand side of (A-25) gives an approximate upper bound on P_{is} for erasing bounded-distance decoders, for loosely packed codes with bounded-distance decoders, and for complete decoders because some received sequences with $t + 1$ or more errors can be corrected and, hence, produce no information-symbol errors.

When there are relatively few sequences in decoding spheres, it is plausible that P_{is} for a reproducing bounded-distance decoder might be accurately estimated

by ignoring undetected errors. Dropping the terms involving $N(l,i)$ in (A-21) and (A-22) and using (A-19) gives

$$P_{is} \geq \sum_{i=t+1}^{n} \binom{n-1}{i-1} P_s^i (1-P_s)^{n-i} \tag{A-26}$$

The virtue of this lower bound as an approximation is its independence of the code weight distribution. The bound is tight when decoding failures are the predominant error mechanism. For cyclic Reed-Solomon codes, numerical examples [40] indicate that the exact P_{is} and the approximate bound are quite close for all values of P_s when $t \geq 3$, a result that is not surprising in view of the paucity of sequences in the decoding spheres for a Reed-Solomon code with $t \geq 3$. A comparison of (A-26) with (A-25) indicates that the latter overestimates P_{is} by a factor of less than $d_m/(t+1)$ as $P_s \to 0$.

A *q-ary symmetric channel* or *uniform discrete channel* is one in which an incorrectly decoded information symbol is equally likely to be any of the remaining $q-1$ symbols in the alphabet. Consider a linear (n,k) block code and a q-ary symmetric channel such that q is a power of 2 and the "channel" refers to the transmission channel plus the decoder. Among the $q-1$ incorrect symbols, a given bit is incorrect in $q/2$ instances. Therefore, the information-bit error probability is

$$P_b = \frac{q}{2(q-1)} P_{is} \tag{A-27}$$

Let r denote the ratio of information bits to transmitted channel symbols. For binary codes, r is the code rate. For block codes with $m = \log_2 q$ information bits per symbol, $r = mk/n$. When coding is used but the information rate is preserved, the duration of a channel symbol is changed relative to that of an information bit. Thus, the energy per received channel symbol is

$$\mathcal{E}_s = r\mathcal{E}_b = \frac{mk}{n}\mathcal{E}_b \tag{A-28}$$

where $\mathcal{E}_b$ is the energy per information bit. When $r < 1$, a code is potentially beneficial if its error-correcting capability is sufficient to overcome the degradation due to the reduction in the energy per received symbol. For the AWGN channel and coherent binary PSK, the classical theory indicates that the symbol error probability at the demodulator output is

$$P_s = Q\left(\sqrt{\frac{2r\mathcal{E}_b}{N_0}}\right) \tag{A-29}$$

where

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp\left(-\frac{y^2}{2}\right) dy = \frac{1}{2} \operatorname{erfc}\left(\frac{x}{\sqrt{2}}\right) \tag{A-30}$$

and erfc( ) is the complementary error function. For noncoherent detection of q-ary orthogonal signals over an AWGN channel, it is shown subsequently that

$$P_s = \sum_{i=1}^{q-1} \frac{(-1)^{i+1}}{i+1} \binom{q-1}{i} \exp\left[-\frac{imr\mathcal{E}_b}{(i+1)N_0}\right] \tag{A-31}$$

which decreases as q increases for sufficiently large values of $\mathcal{E}_b/N_0$. The orthogonality of the signals ensures that at least the transmission channel is q-ary symmetric, and hence (A-27) is at least approximately correct.

If the alphabets of the code symbols and the transmitted channel symbols are the same, then the channel-symbol error probability P_{cs} equals the code-symbol error probability P_s. If not, then the q-ary code symbols may be mapped into q_1-ary channel symbols. If $q = 2^m$ and $q_1 = 2^{m_1}$, then choosing m/m_1 to be an integer is strongly preferred for implementation simplicity. Since any of the channel-symbol errors can cause an error in the corresponding code symbol,

$$P_s = 1 - (1 - P_{cs})^{m/m_1} \tag{A-32}$$

A common application is to map nonbinary code symbols into binary channel symbols ($m_1 = 1$). In this case, (A-27) is no longer valid because the transmission channel plus the decoder is not necessarily q-ary symmetric. Since there is at least one bit error for every symbol error,

$$\frac{P_{is}}{m} \leq P_b \leq \frac{qP_{is}}{2(q-1)} \tag{A-33}$$

This lower bound is tight when P_{cs} is low because then there tends to be a single bit error per code-symbol error before decoding, and the decoder is unlikely to change an information symbol. For coherent binary PSK, (A-29) and (A-32) imply that

$$P_s = 1 - \left[1 - Q\left(\sqrt{\frac{2r\mathcal{E}_b}{N_0}}\right)\right]^m \tag{A-34}$$

Error probabilities for soft-decision decoding

The simplest practical soft-decision decoding entails the use of *erasures* to supplement hard-decision decoding. A symbol is said to be erased when the demodulator instructs the decoder to ignore that symbol during the decoding. An erasure occurs when the demodulator output indicates that the symbol is unreliable. If a code has a minimum distance d_m and a received word is assigned ϵ erasures, then all codewords differ in at least $d_m - \epsilon$ of the unerased symbols. Hence, ν errors can be corrected if $2\nu + 1 \leq d_m - \epsilon$. If d_m or more erasures are assigned, a decoding failure occurs. Let P_ϵ denote the probability of an erasure. For independent symbol errors and erasures, the probability that a received

sequence has i errors and j erasures is $P_s^i P_\epsilon^j (1 - P_s - P_\epsilon)^{n-i-j}$. Therefore, for a bounded-distance decoder,

$$P_w \leq \sum_{j=0}^{n} \sum_{i=i_0}^{n-j} \binom{n}{j} \binom{n-j}{i} P_s^i P_\epsilon^j (1 - P_s - P_\epsilon)^{n-i-j} ,$$
$$i_0 = \max(0, \lceil (d_m - j)/2 \rceil) \qquad \text{(A-35)}$$

where $\lceil x \rceil$ denotes the smallest integer greater than or equal to x. This inequality becomes an equality for an erasing decoder. For the AWGN channel, decoding with optimal erasures provides an insignificant performance improvement relative to hard-decision decoding, but erasures are often effective against fading or sporadic interference. Codes for which errors and erasures decoding is most attractive are those with a relatively large minimum distance such as Reed-Solomon codes.

When soft decisions are made, a number called the *metric* is associated with each possible codeword. The metric is a function of both the codeword and the demodulator output samples. A soft-decision decoder selects the codeword with the largest metric. The information bits are then recovered from this codeword.

Let $\mathbf{y}$ denote the n-dimensional vector of noisy output samples $y_i, i = 1, 2, \ldots, n$, produced by a demodulator that receives a sequence of n symbols. Let $\mathbf{x}_l$ denote the lth codeword vector with symbols $x_{li}, i = 1, 2, \ldots, n$. Let $f(\mathbf{y}|\mathbf{x}_l)$ denote the *likelihood function*, which is the conditional probability density function of $\mathbf{y}$ given that $\mathbf{x}_l$ was transmitted. The *maximum-likelihood decoder* finds the value of l, $1 \leq l \leq q^k$, for which the likelihood function is largest. If this value is l_0, the decoder decides that codeword l_0 was transmitted. Any monotonically increasing function of $f(\mathbf{y}|\mathbf{x}_l)$ may serve as the metric of a maximum-likelihood decoder. A convenient choice is often proportional to the logarithm of $f(\mathbf{y}|\mathbf{x}_l)$, which is called the *log-likelihood function*. If the demodulator outputs are statistically independent and a single output corresponds to each code symbol, then the log-likelihood function for each of the q^k possible codeword is

$$\ln f(\mathbf{y}|\mathbf{x}_l) = \sum_{i=1}^{n} \ln f(y_i|x_{li}) , \qquad l = 1, 2, \ldots, q^k \qquad \text{(A-36)}$$

where $f(y_i|x_{li})$ is the conditional probability density function of y_i given the value of x_{li}.

For binary PSK communications over the AWGN channel, if codeword l is transmitted, then the received signal representing symbol i is

$$r_i(t) = \sqrt{2\mathcal{E}_s} x_{li} \psi(t) \cos 2\pi f_c t + n_i(t) , \quad 0 \leq t \leq T_s, \quad i = 1, 2, \ldots, n \quad \text{(A-37)}$$

where $\mathcal{E}_s$ is the symbol energy, T_s is the symbol duration, f_c is the carrier frequency, $x_{li} = +1$ when binary symbol i is a 1 and $x_{li} = -1$ when binary

symbol i is a 0, $\psi(t)$ is the unit-energy symbol waveform, and $n_i(t)$ is independent, zero-mean, white Gaussian noise. Since $\psi(t)$ has unit energy and vanishes outside $[0, T_s]$,

$$\int_0^{T_s} |\psi(t)|^2 dt = 1 \tag{A-38}$$

For coherent demodulation, a frequency translation to baseband is provided by multiplying $r_i(t)$ by $\cos 2\pi f_c t$. After discarding a negligible integral, we find that the matched-filter demodulator, which is matched to $\psi(t)$, produces the output samples

$$y_i = \sqrt{\mathcal{E}_s/2}x_{li} + \int_0^{T_s} n_i(t)\psi(t)\cos 2\pi f_c t \, dt \ , \quad i = 1, 2, \ldots, n \tag{A-39}$$

These outputs provide sufficient statistics because $\psi(t)\cos 2\pi f_c t$ is the sole basis function for the signal space. Since $n_i(t)$ is statistically independent of $n_k(t)$ when $i \neq k$, the $\{y_i\}$ are statistically independent.

The autocorrelation of each white noise process is

$$E[n_i(t)n_i(t+\tau)] = \frac{N_{0i}}{2}\delta(\tau) \ , \quad i = 1, 2, \ldots, n \tag{A-40}$$

where $N_{0i}/2$ is the two-sided power spectral density of $n_i(t)$ and $\delta(\tau)$ is the Dirac delta function. A straightforward calculation using (A-40) and assuming that the spectrum of $\psi(t)$ is confined to $|f| < f_c$ indicates that the variance of the noise term of (A-39) is $N_{0i}/4$. Therefore, the conditional probability density function of y_i given that x_{li} was transmitted is

$$f(y_i|x_{li}) = \frac{1}{\sqrt{\pi N_{0i}/2}} \exp\left[-\frac{(y_i - \sqrt{\mathcal{E}_s/2}x_{li})^2}{N_{0i}/2}\right] \ , \quad i = 1, 2, \ldots, n \tag{A-41}$$

Since y_i^2 and $x_{li}^2 = 1$ are independent of the codeword l, terms involving these quantities may be discarded in the log-likelihood function of (A-36). Therefore, the maximum-likelihood metric is

$$U(l) = \sum_{i=1}^{n} \frac{x_{li}y_i}{N_{0i}} \ , \quad l = 1, 2, \ldots, 2^k \tag{A-42}$$

which requires knowledge of $N_{0i}, i = 1, 2, \ldots, n$.

If each $N_{0i} = N_0$, a constant, then this constant is irrelevant, and the maximum-likelihood metric is

$$U(l) = \sum_{i=1}^{n} x_{li}y_i \ , \quad l = 1, 2, \ldots, 2^k \tag{A-43}$$

Let $P_2(\delta)$ denote the probability that the metric for an incorrect codeword at distance δ from the correct codeword exceeds the metric for the correct

codeword. After reordering the samples $\{y_i\}$, the difference between the metrics for the correct codeword and the incorrect one may be expressed as

$$D(\delta) = \sum_{i=1}^{\delta} (x_{1i} - x_{2i}) y_i = 2 \sum_{i=1}^{\delta} x_{li} y_i \tag{A-44}$$

where the sum includes only the δ terms that differ, x_{1i} refers to the correct codeword, x_{2i} refers to the incorrect codeword, and $x_{2i} = -x_{1i}$. Then $P_2(\delta)$ is the probability that $D(\delta) < 0$. Since each of its terms is independent, $D(\delta)$ has a Gaussian distribution. A straightforward calculation using (A-41) and $\mathcal{E}_s = r\mathcal{E}_b$ yields

$$P_2(\delta) = Q\left(\sqrt{\frac{2\delta r \mathcal{E}_b}{N_0}}\right) \tag{A-45}$$

which reduces to (A-29) when a single symbol is considered and $\delta = 1$.

A fundamental property of a probability, called *countable subadditivity*, is that the probability of a finite or countable union of events B_n, $n = 1, 2, \ldots$, satisfies

$$P\left(\cup_n B_n\right) \leq \sum_n P(B_n) \tag{A-46}$$

In communication theory, a bound obtained from this inequality is called a *union bound*. To determine P_w for linear block codes, it suffices to assume that the all-zero codeword was transmitted. The union bound and the relation between weights and distances imply that P_w for soft-decision decoding satisfies

$$P_w \leq \sum_{l=d_m}^{n} A_l P_2(l) \tag{A-47}$$

Let β_l denote the total information-symbol weight of the codewords of weight l. The union bound and (A-16) imply that

$$P_{is} \leq \sum_{l=d_m}^{n} \frac{\beta_l}{k} P_2(l) \tag{A-48}$$

To determine β_l for any cyclic (n, k) code, consider the set S_l of A_l codewords of weight l. The total weight of all the codewords in S_l is $A_T = lA_l$. Let α and β denote any two fixed positions in the codewords. By definition, any cyclic shift of a codeword produces another codeword of the same weight. Therefore, for every codeword in S_l that has a zero in α, there is some codeword in S_l that results from a cyclic shift of that codeword and has a zero in β. Thus, among the codewords of S_l, the total weight of all the symbols in a fixed position is the same regardless of the position and is equal to A_T/n. The total weight of all the information symbols in S_l is $\beta_l = kA_T/n = klA_l/n$. Therefore,

$$P_{is} \leq \sum_{l=d_m}^{n} \frac{l}{n} A_l P_2(l) \tag{A-49}$$

Although it is theoretically advantageous, optimal soft-decision decoding cannot be efficiently implemented except for very short block codes, primarily because the number of codewords for which the metrics must be computed is prohibitively large. Approximate maximum-likelihood decoding algorithms are available. The *Chase algorithm* [38] generates a small set of candidate codewords that will almost always include the codeword with the largest metric. Test patterns are generated by first making hard decisions on each of the received symbols and then altering the least reliable symbols, which are determined from the demodulator outputs given by (A-39). Hard-decision decoding of each test pattern and the discarding of decoding failures generates the candidate codewords. The decoder selects the candidate codeword with the largest metric.

For digital computations to be performed, soft-decision information must be quantized. Two levels of quantization correspond to hard decisions. More than two levels require analog-to-digital conversion of the demodulator output samples. Because the optimal location of the levels is a function of the signal, thermal noise, and interference powers, automatic gain control is often necessary. For the AWGN channel, it is found that an eight-level quantization represented by three bits and a uniform spacing between threshold levels cause no more than a few tenths of a decibel loss relative to what could theoretically be achieved with unquantized analog voltages or infinitely fine quantization.

Calculations for specific communication systems and codes operating over the AWGN channel have shown that approximately 2 dB of additional signal power is required for a hard-decision receiver to produce the same information-bit error rate as the corresponding optimal soft-decision receiver. However, soft-decision receivers are much more complex to implement and may be too slow for the processing of high information rates. For a given level of implementation complexity, hard-decision decoders can accommodate much longer block codes, thereby at least partially overcoming the inherent advantage of soft-decision decoders. In practice, soft-decision decoding other than erasures is seldom used with block codes of length greater than 50.

Performance examples

Figure A-2 depicts the information-bit error probability $P_b = P_{is}$ versus $\mathcal{E}_b/N_0$ for various binary block codes with coherent PSK over the AWGN channel. Equation (A-25) is used to compute P_b for the Golay (23,12) code with hard decisions. Since the packing density D_p is small for these codes, (A-26) is used for the BCH (63,36) code, which corrects $t = 5$ errors, and the BCH (127,64) code, which corrects $t = 10$ errors. Equation (A-29) is used for P_s. Inequality (A-49) and Table A-2 are used to compute the upper bound on $P_b = P_{is}$ for the Golay (23,12) code with optimal soft decisions. The curves illustrate the power of the soft-decision decoding. For the Golay (23,12) code, soft-decision decoding requires approximately 2 dB less $\mathcal{E}_b/N_0$ to achieve $P_b = 10^{-5}$ than hard-decision decoding. Only when $P_b < 10^{-5}$ does the BCH (127,64) begin to outperform the

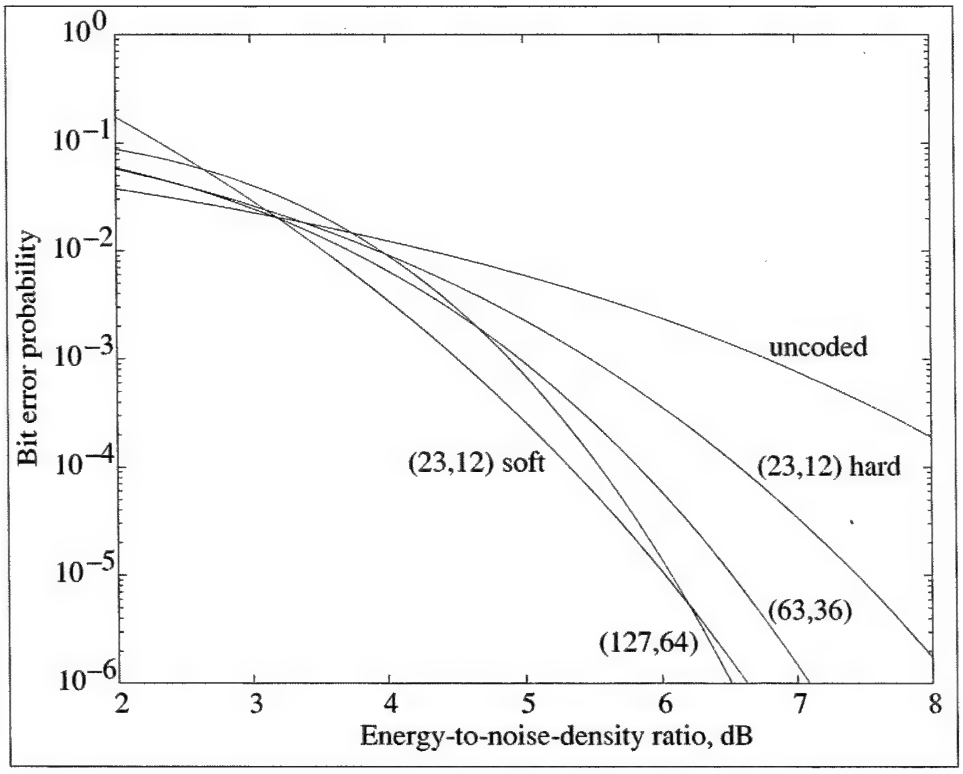


Figure A-2. Information-bit error probability for binary block (n, k) codes and coherent PSK.

Golay (23,12) code with soft decisions. If $\mathcal{E}_b/N_0 \leq 3$ dB, an uncoded system with coherent PSK provides a lower P_b than a similar system that uses one of the block codes of the figure.

Figure A-3 illustrates the performance of loosely packed Reed-Solomon codes with hard-decision decoding over the AWGN channel. The lower bound in (A-26) is used to compute the approximate information-bit error probabilities for binary channel symbols with coherent PSK and for nonbinary channel symbols with noncoherent MFSK. For the nonbinary channel symbols, (A-27) and (A-31) are used. For the binary channel symbols, (A-34) and the lower bound in (A-33) are used. For the chosen values of n, the best performance at $P_b = 10^{-5}$ is obtained if the code rate is $k/n \approx 3/4$. Further gains result from increasing n and hence the implementation complexity. Although the figure indicates the performance advantage of Reed-Solomon codes with MFSK, there is a major bandwidth penalty. Let B denote the bandwidth required for an uncoded binary PSK signal. If the same data rate is accommodated by using uncoded binary FSK, the required bandwidth for demodulation with envelope detectors is approximately $2B$. For uncoded MFSK using $q = 2^m$ frequencies, the required bandwidth is $2^m B/m$ because each symbol represents m bits. If a Reed-Solomon (n, k) code is used with MFSK, the required bandwidth becomes $2^m nB/mk$.

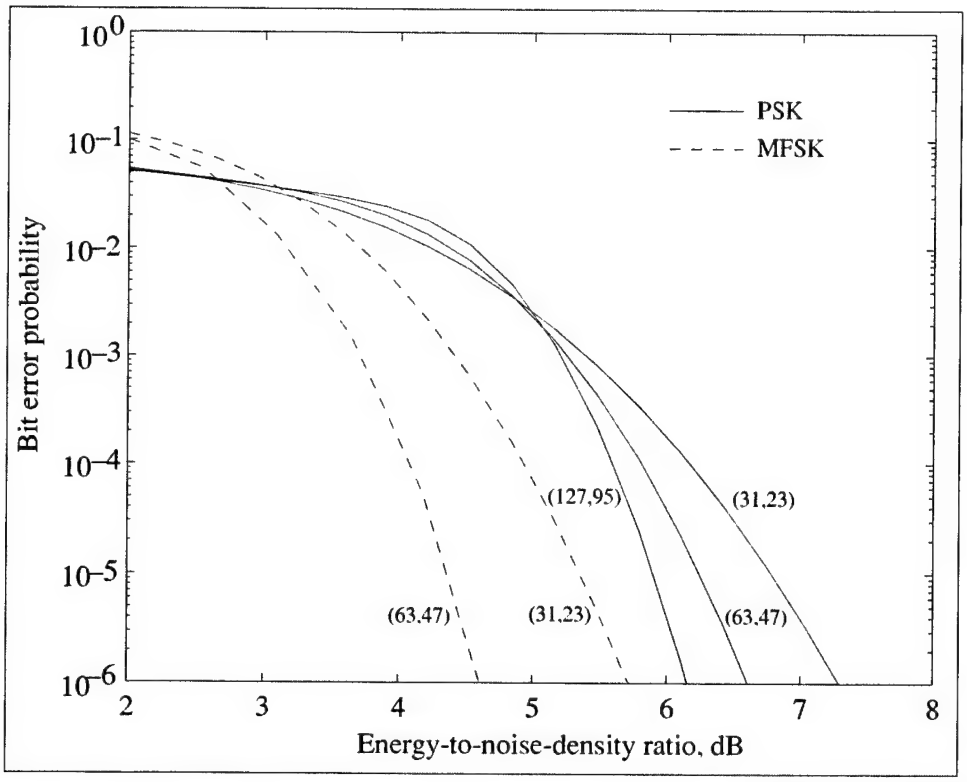


Figure A-3. Information-bit error probability for Reed-Solomon (n, k) codes. Modulation is coherent PSK or noncoherent MFSK.

Metrics and error probabilities for MFSK

For q-ary orthogonal symbol waveforms, $s_1(t), s_2(t), \ldots, s_q(t)$, q matched filters are needed, and the observation vector is $\mathbf{y} = [\mathbf{y}_1\ \mathbf{y}_2 \ldots\ \mathbf{y}_q]$, where each $\mathbf{y}_k$ is an n-dimensional row vector of matched-filter output samples for filter k with components $y_{ki}, i = 1, 2, \ldots, n$. Suppose that symbol i of codeword l uses unit-energy waveform $s_\nu(t)$, where ν is a function of i and l. If codeword l is transmitted over the AWGN channel, the received signal for symbol i can be expressed in complex notation as

$$r_i(t) = \mathrm{Re}\left[\sqrt{2\mathcal{E}_s}s_\nu(t)e^{j2\pi f_c t+\theta_i}\right] + n_i(t), \quad 0 \le t \le T_s\,, \quad i = 1, 2, \ldots, n \tag{A-50}$$

where $n_i(t)$ is independent, zero-mean, white Gaussian noise with two-sided power spectral density $N_{0i}/2$, f_c is the carrier frequency, and θ_i is the phase. Since the symbol energy for all the waveforms is unity,

$$\int_0^{T_s} |s_k(t)|^2 dt = 1\,, \quad k = 1, 2, \ldots, q \tag{A-51}$$

The orthogonality of symbol waveforms implies that

$$\int_0^{T_s} s_m(t)s_n^*(t)dt = 0\,, \quad m \neq n \tag{A-52}$$

A frequency translation to baseband is followed by matched filtering. Matched-filter k, which is matched to $s_k(t)$, produces the output samples

$$y_{ki} = \int_0^{T_s} r_i(t) e^{-j2\pi f_c t} s_k^*(t) dt \ , \quad i = 1, 2, \ldots, n \ , \quad k = 1, 2, \ldots, q \qquad \text{(A-53)}$$

The substitution of (A-50) into (A-53), (A-52), and the assumption that each of the $\{s_k(t)\}$ has a spectrum confined to $|f| < f_c$ yields

$$y_{ki} = \sqrt{\mathcal{E}_s/2} e^{j\theta_i} \delta_{k\nu} + \int_0^{T_s} n_i(t) e^{-j2\pi f_c t} s_k^*(t) dt \qquad \text{(A-54)}$$

where $\delta_{k\nu} = 1$ if $k = \nu$ and $\delta_{k\nu} = 0$ otherwise.

Equation (A-40) implies that the real and imaginary components of the noise term in (A-54), which are jointly Gaussian, are uncorrelated and, hence, independent and have the same variance $N_{0i}/4$. Since the density of a complex-valued random variable is defined to be the joint density of its real and imaginary parts, the conditional probability density function of y_{ki} given θ_i and $\delta_{k\nu}$ is

$$f\left(y_{ki} \,|\delta_{k\nu}, \theta_i\right) = \frac{1}{\pi N_{0i}/2} \exp\left(-\frac{\left| y_{ki} - \sqrt{\mathcal{E}_s/2} e^{j\theta_i} \delta_{k\nu} \right|^2}{N_{0i}/2} \right) \qquad \text{(A-55)}$$

The independence of the $\{n_i(t)\}$, the orthogonality condition (A-52), and the white Gaussian character of the noise ensure the conditional independence of the $\{y_{ki}\}$ for all values of k and i.

For coherent signals, the $\{\theta_i\}$ are tracked by the phase synchronization system and thus ideally may be set to zero. The likelihood function is the product of qn densities given by (A-55). Forming the log-likelihood function with the $\{\theta_i\}$ set to zero, and eliminating irrelevant terms that are independent of l, we obtain the maximum-likelihood metric

$$U(l) = \sum_{i=1}^{n} \frac{\mathrm{Re}(V_{li})}{N_{0i}} \qquad \text{(A-56)}$$

where $V_{li} = y_{\nu i}$ is the sampled output of the filter matched to $s_\nu(t)$, the signal representing symbol i of codeword l. If each $N_{0i} = N_0$, then the maximum-likelihood metric is

$$U(l) = \sum_{i=1}^{n} \mathrm{Re}\left(V_{li}\right) \qquad \text{(A-57)}$$

and the common value N_0 does not need to be known to apply this metric.

For noncoherent signals, it is assumed that each θ_i is independent and uniformly distributed over $[0, 2\pi)$, which preserves the conditional independence of the

$\{y_{ki}\}$ given the $\{\delta_{k\nu}\}$. Expanding the argument of the exponential function in (A-55), expressing y_{ki} in polar form, and integrating over θ_i, we obtain the conditional probability density function

$$f\left(y_{ki}\,|\delta_{k\nu}\right) = \frac{1}{\pi N_{0i}/2} \exp\left[-\frac{|y_{ki}|^2 + \mathcal{E}_s\delta_{k\nu}/2}{N_{0i}/2}\right] I_0\left(\frac{\sqrt{8\mathcal{E}_s}\,|y_{ki}|\,\delta_{k\nu}}{N_{0i}}\right) \quad \text{(A-58)}$$

where $I_0(\,)$ is the modified Bessel function of the first kind and order zero, and

$$I_0(x) = \frac{1}{2\pi}\int_0^{2\pi} \exp(x\cos u)du \quad \text{(A-59)}$$

Let $R_{li} = |y_{\nu i}|$ denote the sampled envelope produced by the filter matched to $s_\nu(t)$, the signal representing symbol i of codeword l. We form the log-likelihood function and eliminate terms and factors that do not depend on the codeword l, thereby obtaining the maximum-likelihood metric

$$U(l) = \sum_{i=1}^{n} \ln I_0\left(\frac{\sqrt{8\mathcal{E}_s}R_{li}}{N_{0i}}\right) \quad \text{(A-60)}$$

If each $N_{0i} = N_0$, then the maximum-likelihood metric is

$$U(l) = \sum_{i=1}^{n} \ln I_0\left(\frac{\sqrt{8\mathcal{E}_s}R_{li}}{N_0}\right) \quad \text{(A-61)}$$

and $\sqrt{\mathcal{E}_s}/N_0$ must be known to apply this metric.

From the series representation of $I_0(x)$, it follows that

$$I_0(x) \le \exp\left(\frac{x^2}{4}\right) \quad \text{(A-62)}$$

From (A-59), we obtain

$$I_0(x) \le \exp(x)\,, \quad x \ge 0 \quad \text{(A-63)}$$

The upper bound in (A-62) is tighter for $0 \le x < 2$, while the upper bound in (A-63) is tighter for $2 < x < \infty$. If we assume that R_{li}/N_{0i} is often less than 2, then the approximation of $I_0(x)$ by $\exp(x^2/4)$ is reasonable. Substitution into (A-60) and dropping an irrelevant constant gives the metric

$$U(l) = \sum_{i=1}^{n} \frac{R_{li}^2}{N_{0i}^2} \quad \text{(A-64)}$$

If each $N_{0i} = N_0$, then the value of N_0 is irrelevant, and we obtain the *Rayleigh metric*

$$U(l) = \sum_{i=1}^{n} R_{li}^2 \quad \text{(A-65)}$$

which is the maximum-likelihood metric for a Rayleigh fading channel. Similarly, (A-63) can be used to obtain suboptimal metrics suitable for large values of R_{li}/N_{0i}.

To determine the maximum-likelihood metric for making a hard decision on each symbol, we set $n = 1$ and drop the subscript i in (A-60). We find that the maximum-likelihood symbol metric is $\ln [I_0(\sqrt{8\mathcal{E}_s}R_l/N_0)]$, where the index l ranges over the symbol alphabet. However, since this function increases monotonically and $\sqrt{8\mathcal{E}_s}/N_0$ is a constant, the optimal symbol metrics or decision variables are R_l or R_l^2 for $l = 1, 2, \ldots, q$.

For noncoherent MFSK, matched-filter l is matched to the unit-energy waveform $s_l(t) = A\exp(j2\pi f_l t)$, $0 \le t \le T_s$, where $A = 1/\sqrt{T_s}$. If $r(t)$ is the received signal, a frequency translation to baseband and a parallel set of matched filters and envelope detectors provide the decision variables

$$R_l^2 = A^2 \left| \int_0^{T_s} r(t) e^{-j2\pi f_c t} e^{-j2\pi f_l t} dt \right|^2 \tag{A-66}$$

The orthogonality condition (A-52) is satisfied if the adjacent frequencies are separated by k/T_s, where k is a nonzero integer. Expanding (A-66), we obtain

$$R_l^2 = R_{lc}^2 + R_{ls}^2 \tag{A-67}$$

$$R_{lc} = A \int_0^{T_s} r(t) \cos [2\pi(f_c + f_l)t] \, dt \tag{A-68}$$

$$R_{ls} = A \int_0^{T_s} r(t) \sin [2\pi(f_c + f_l)t] \, dt \quad . \tag{A-69}$$

These equations imply the correlator structure depicted in Figure A-4, where the irrelevant constant A has been omitted. The comparator decides what symbol was transmitted by observing which comparator input is the largest.

To derive an alternative implementation, we observe that when the waveform is $s_l(t) = A\cos 2\pi(f_c + f_l)t$, $0 \le t \le T_s$, the impulse response of a filter matched to it is $A\cos 2\pi(f_c + f_l)(T_s - t)$, $0 \le t \le T_s$. Therefore, the matched-filter output at time t is

$$\begin{aligned} y_l(t) &= A \int_0^t r(\tau) \cos [2\pi(f_c + f_l)(\tau - t + T_s)] \, d\tau \\ &= A \left\{ \int_0^t r(\tau) \cos [2\pi(f_c + f_l)\tau] \, d\tau \right\} \cos [2\pi(f_c + f_l)(t - T_s)] \\ &\quad + A \left\{ \int_0^t r(\tau) \sin [2\pi(f_c + f_l)\tau] \, d\tau \right\} \sin [2\pi(f_c + f_l)(t - T_s)] \\ &= AR_l(t) \cos [2\pi(f_c + f_l)(t - T_s) + \phi(t)] \ , \quad 0 \le t \le T_s \end{aligned} \tag{A-70}$$

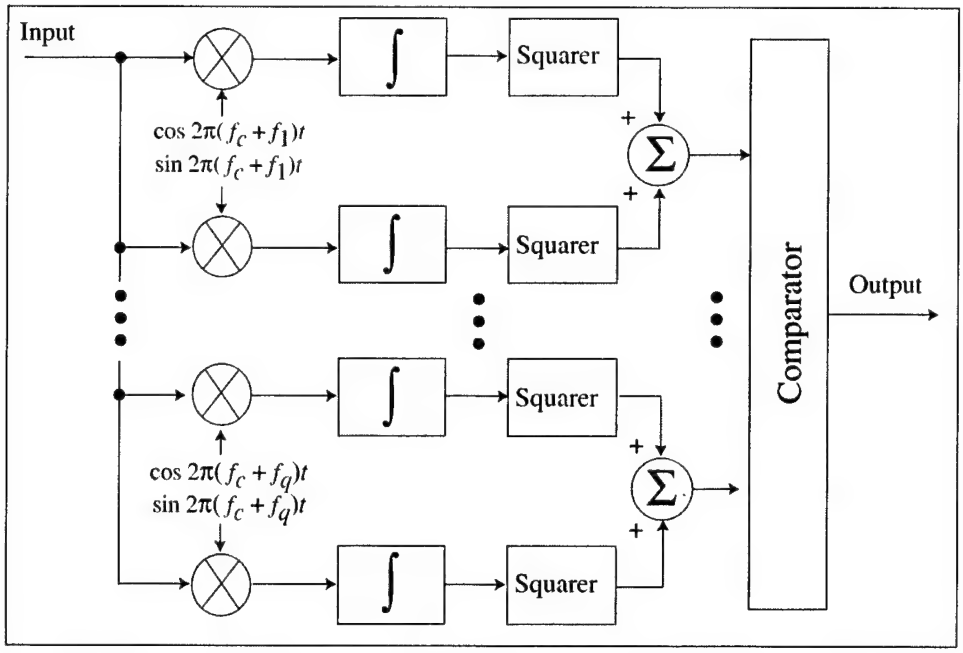


Figure A-4. Noncoherent MFSK receiver using correlators.

where the envelope is

$$R_l(t) = A\left\{ \left[\int_0^t r(\tau)\cos\left[2\pi(f_c + f_l)\tau\right] d\tau\right]^2 + \left[\int_0^t r(\tau)\sin\left[2\pi\left(f_c + f_l\right)\tau\right] d\tau\right]^2 \right\}^{1/2} \tag{A-71}$$

Since $R_l(T_s) = R_l$ given by (A-67), we obtain the receiver structure depicted in Figure A-5, where the irrelevant constant A has been omitted. A practical envelope detector consists of a peak detector followed by a lowpass filter.

To derive the symbol error probability for equally likely MFSK symbols, we assume that the signal $s_1(t)$ was transmitted over the AWGN channel. The received signal has the form $r(t) = \sqrt{2\mathcal{E}_s/T_s}\cos\left[2\pi\left(f_c + f_1\right)t + \theta\right] + n(t)$, $0 \le t \le T_s$. Since $n(t)$ is white,

$$E[n(t)n(t+\tau)] = \frac{N_0}{2}\delta(\tau) \tag{A-72}$$

Using the orthogonality of the symbol waveforms and assuming that $f_c + f_l >> 1/T_s$ in (A-68) and (A-69), we obtain

$$E[R_{lc}] = \sqrt{\mathcal{E}_s/2}\cos\theta\,, \quad E[R_{ls}] = \sqrt{\mathcal{E}_s/2}\sin\theta \tag{A-73}$$

$$E[R_{lc}] = E[R_{ls}] = 0\,, \quad l = 2, \ldots, q \tag{A-74}$$

$$\mathrm{var}(R_{lc}) = \mathrm{var}(R_{ls}) = N_0/4\,, \quad l = 1, 2, \ldots, q \tag{A-75}$$

Since $n(t)$ is Gaussian, R_{lc} and R_{ls} are jointly Gaussian. Since the covariance of R_{lc} and R_{ls} is zero, they are mutually statistically independent. Therefore, the

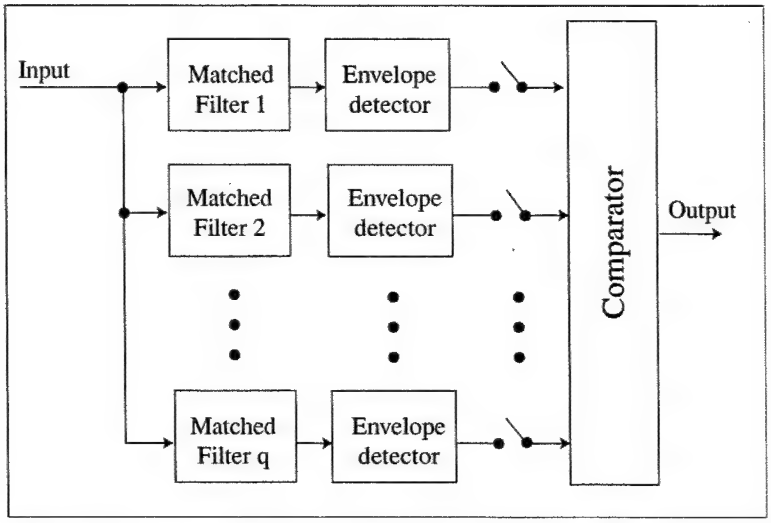


Figure A-5. Noncoherent MFSK receiver with passband matched filters.

joint probability density function of R_{lc} and R_{ls} is

$$g_1(r_{lc}, r_{ls}) = \frac{1}{\pi N_0/2} \exp\left[-\frac{(r_{lc} - m_{lc})^2 + (r_{ls} - m_{ls})^2}{N_0/2}\right] \tag{A-76}$$

where $m_{lc} = E[R_{lc}]$ and $m_{ls} = E[R_{ls}]$.

Let R_l and Θ_l be implicitly defined by $R_{lc} = R_l \cos\Theta_l$ and $R_{ls} = R_l \sin\Theta_l$. Using the Jacobian of the transformation, we find that the joint density of R_l and Θ_l is

$$g_2(r, \theta) = \frac{2r}{\pi N_0} \exp\left[-\frac{r^2 - 2rm_{lc}\cos\theta - 2rm_{ls}\sin\theta + m_{lc}^2 + m_{ls}^2}{N_0/2}\right]$$
$$r \geq 0, \quad |\theta| \leq \pi \tag{A-77}$$

The density of the envelope R_l is obtained by integration of (A-77) over θ. Using trigonometry and (A-59), we obtain the density

$$g_3(r) = \frac{4r}{N_0} \exp\left(-\frac{r^2 + m_{lc}^2 + m_{ls}^2}{N_0/2}\right) I_0\left(\frac{4r\sqrt{m_{lc}^2 + m_{ls}^2}}{N_0}\right) u(r) \tag{A-78}$$

where $u(r) = 1$ if $r \geq 0$, and $u(r) = 0$ if $r < 0$. Substituting (A-73), we obtain the densities for the R_l, $l = 1, 2, \ldots, q$:

$$f_1(r) = \frac{4r}{N_0} \exp\left(-\frac{r^2 + \mathcal{E}_{s/2}}{N_0/2}\right) I_0\left(\frac{\sqrt{8\mathcal{E}_s}r}{N_0}\right) u(r) \tag{A-79}$$

$$f_l(r) = \frac{4r}{N_0} \exp\left(-\frac{r^2}{N_0/2}\right) u(r), \quad l = 2, \ldots, q \tag{A-80}$$

The orthogonality of the symbol waveforms and (A-72) imply that the random variables $\{R_l\}$ are independent. A symbol error occurs when $s_1(t)$ was

transmitted if R_1 is not the largest of the $\{R_l\}$. Since the $\{R_l\}$ are identically distributed for $l = 2, \cdots, q$, the probability of a symbol error when $s_1(t)$ was transmitted is

$$P_s = 1 - \int_0^\infty \left[\int_0^r f_2(y)dy\right]^{q-1} f_1(r)dr \tag{A-81}$$

Substituting (A-80) into the inner integral gives

$$\int_0^r f_2(y)dy = 1 - \exp\left(-\frac{r^2}{N_0/2}\right) \tag{A-82}$$

Expressing the $(q-1)$th power of this result as a binomial expansion and then substituting into (A-81), the remaining integration may be done by using the fact that for $\lambda > 0$,

$$\int_0^\infty r \exp\left(-\frac{r^2}{2b^2}\right) I_0\left(\frac{r\sqrt{\lambda}}{b^2}\right) dr = b^2 \exp\left(\frac{\lambda}{2b^2}\right) \tag{A-83}$$

which follows from the fact that the density in (A-79) must integrate to unity. The final result is the symbol error probability for noncoherent MFSK over the AWGN channel:

$$P_s = \sum_{i=1}^{q-1} \frac{(-1)^{i+1}}{i+1} \binom{q-1}{i} \exp\left[-\frac{i\mathcal{E}_s}{(i+1)N_0}\right] \tag{A-84}$$

A generalization for binary FSK is easily derived. Let $N_{01}/2$ and $N_{02}/2$ denote the two-sided power spectral densities of the white Gaussian noise in the filter matched to the transmitted signal and the other matched filter, respectively. Changing N_0 to N_{01} in (A-79) and N_0 to N_{02} in (A-80), substituting into (A-81) with $q = 2$, and evaluating the integrals by using (A-83), we obtain

$$P_s = \frac{N_{02}}{N_{01} + N_{02}} \exp\left(-\frac{\mathcal{E}_s}{N_{01} + N_{02}}\right) \tag{A-85}$$

Therefore, if it is equally likely that either $N_{01}/2$ or $N_{02}/2$ is the noise density in the filter matched to a transmitted symbol, then interchanging the roles of N_{01} and N_{02} in (A-85) and averaging over the two possibilities gives the symbol error probability for noncoherent FSK over the AWGN channel:

$$P_s = \frac{1}{2} \exp\left(-\frac{\mathcal{E}_s}{N_{01} + N_{02}}\right) \tag{A-86}$$

which depends only on the average noise power spectral density of the two matched filters, not on the individual levels N_{01} and N_{02}. If $N_{01} = N_{02} = N_0$, then (A-86) reduces to the classical formula

$$P_s = \frac{1}{2} \exp\left(-\frac{\mathcal{E}_s}{2N_0}\right) \tag{A-87}$$

When frequency-nonselective fading occurs, the symbol energy may be expressed as $\mathcal{E}_s\alpha^2$, where $\mathcal{E}_s$ represents the average energy and α is a random variable with $E[\alpha^2] = 1$. For Ricean fading, the probability density function of α is [15]

$$f_\alpha(r) = 2(\kappa+1)r\exp\{-\kappa-(\kappa+1)r^2\}I_0(\sqrt{\kappa(\kappa+1)}2r)u(r) \qquad \text{(A-88)}$$

where κ is Rice factor. Replacing $\mathcal{E}_s$ by $\mathcal{E}_s\alpha^2$ in (A-84), an integration over the density (A-88) and the use of (A-83) yield

$$P_s = \sum_{i=1}^{q-1}(-1)^{i+1}\binom{q-1}{i}\frac{\kappa+1}{\kappa+1+(\kappa+1+\mathcal{E}_s/N_0)i} \exp\left[-\frac{\kappa i\mathcal{E}_s/N_0}{\kappa+1+(\kappa+1+\mathcal{E}_s/N_0)i}\right] \qquad \text{(A-89)}$$

For binary FSK,

$$P_s = \frac{\kappa+1}{2(\kappa+1)+\mathcal{E}_s/N_0}\exp\left[-\frac{\kappa\mathcal{E}_s/N_0}{2(\kappa+1)+\mathcal{E}_s/N_0}\right] \qquad \text{(A-90)}$$

Chernoff Bound

The *moment generating function* of the random variable X with distribution function $F(\alpha)$ is defined as

$$M(s) = E\left[e^{sX}\right] = \int_{-\infty}^{\infty}\exp(sx)dF(x) \qquad \text{(A-91)}$$

for all real s for which the integral is finite. For all nonnegative s, the probability that $X \geq 0$ is

$$P[X \geq 0] = \int_0^{\infty} dF(x) \leq \int_0^{\infty}\exp(sx)dF(x) \qquad \text{(A-92)}$$

Comparing (A-91) and (A-92), we conclude that

$$P[X \geq 0] \leq M(s), \quad 0 \leq s < s_1 \qquad \text{(A-93)}$$

where s_1 is the upper limit of an open interval in which $M(s)$ is defined. To make this bound as tight as possible, we choose the value of s that minimizes $M(s)$. Therefore,

$$P[X \geq 0] \leq \min_{0\leq s<s_1} M(s) \qquad \text{(A-94)}$$

The right-hand side of this inequality is called the *Chernoff bound*. It is potentially useful if it can be more easily evaluated than $P[X \geq 0]$. From (A-94) and (A-91), we obtain

$$P[X \geq b] \leq \min_{0\leq s<s_1} M(s)\exp(-sb) \qquad \text{(A-95)}$$

Since the moment generating function is finite in some neighborhood of $s = 0$, we may differentiate under the integral sign in (A-91) to obtain the derivative of $M(s)$. The result is

$$M'(s) = \int_{-\infty}^{\infty} x \exp(sx) dF(x) \tag{A-96}$$

which implies that $M'(0) = E[X]$. Differentiating (A-96) gives the second derivative

$$M''(s) = \int_{-\infty}^{\infty} x^2 \exp(sx) dF(x) \tag{A-97}$$

This equation shows that $M''(s) \geq 0$, which implies that $M(s)$ is convex in its interval of definition. Suppose that

$$E(X) < 0\,, \quad P[X > 0] > 0 \tag{A-98}$$

The first inequality implies that $M'(0) < 0$, and the second inequality implies that $M(s) \to \infty$ as $s \to \infty$. Thus, since $M(0) = 1$ and $M(s)$ is convex, it has its minimum value less than unity at some positive $s = s_0$. We conclude that (A-98) is sufficient to ensure that the Chernoff bound is less than unity and $s_0 > 0$.

The Chernoff bound can be tightened if X has a density function $f(x)$ such that

$$f(-x) \geq f(x)\,, \quad x \geq 0 \tag{A-99}$$

For s in A, where A is the open interval over which $M(s)$ is finite, (A-91) implies that

$$\begin{aligned} M(s) &= \int_0^{\infty} \exp(sx) f(x) dx + \int_{-\infty}^{0} exp(sx) f(x) dx \\ &\geq \int_0^{\infty} [\exp(sx) + \exp(-sx)]\, f(x) dx = \int_0^{\infty} 2\cosh(sx) f(x) dx \\ &\geq 2 \int_0^{\infty} f(x) dx = 2P[X \geq 0] \end{aligned} \tag{A-100}$$

Thus, we obtain

$$P[X \geq 0] \leq \frac{1}{2} \min_{s \in A} M(s) \tag{A-101}$$

In this version of the Chernoff bound, the minimum value s_0 is not required to be nonnegative. However, if (A-98) holds, then the bound is less than 1/2, $s_0 > 0$, and

$$P[X \geq 0] \leq \frac{1}{2} \min_{0 < s < s_1} M(s) \tag{A-102}$$

In soft-decision decoding, the encoded sequence or codeword with the largest associated metric is converted into the decoded output. Let $U(j)$ denote the value of the metric associated with sequence j of length L. Consider additive metrics having the form

$$U(j) = \sum_{i=1}^{L} m(j,i) \tag{A-103}$$

where $m(j,i)$ is the *symbol metric* associated with symbol i of the encoded sequence. Let $j = 1$ label the correct sequence and $j = 2$ label an incorrect one. By suitably relabeling the l symbol metrics that may differ for the two sequences, we obtain

$$\begin{aligned} P_2(l) &\leq P[U(2) \geq U(1)] \\ &= P\left[\sum_{i=1}^{l} [m(2,i) - m(1,i)] \geq 0\right] \end{aligned} \tag{A-104}$$

where the inequality results because $U(2) = U(1)$ does not necessarily cause an error if it occurs. In all practical cases, (A-98) is satisfied for the random variable $X = U(2) - U(1)$. Therefore, the Chernoff bound implies that

$$P_2(l) \leq \alpha \min_{0<s<s_1} E\left[\exp\left\{s \sum_{i=1}^{l} [m(2,i) - m(1,i)]\right\}\right] \tag{A-105}$$

where s_1 is the upper limit of the interval over which the expected value is defined. If (A-99) is satisfied, then $\alpha = 1/2$; otherwise, $\alpha = 1$.

If $m(2,i) - m(1,i)$, $i = 1, 2, \ldots, l$, are independent, identically distributed random variables and we define

$$Z = \min_{0<s<s_1} E[\exp\{s[m(2,i) - m(1,i)]\}] \tag{A-106}$$

then the Chernoff bound is given by

$$P_2(l) \leq \alpha Z^l \tag{A-107}$$

This bound is often much simpler to compute than the exact $P_2(l)$. As l increases, the central-limit theorem implies that the distribution of $X = U(2) - U(1)$ approximates the Gaussian distribution. Thus, for large enough l, (A-99) is satisfied when $E[X] < 0$, and we set $\alpha = 1/2$ in (A-107). For small l, (A-99) may be difficult to establish mathematically, but is often intuitively clear; if not, setting $\alpha = 1$ in (A-107) is always valid.

These results can be applied to hard-decision decoding, which can be regarded as a special case of soft-decision decoding with the following symbol metric. If symbol i of a candidate sequence j agrees with the corresponding detected

symbol at the demodulator output, then $m(j,i) = 1$; otherwise $m(j,i) = 0$. Therefore, $m(2,i) - m(1,i)$ in (A-106) is equal to +1 with probability P_s and -1 with probability $(1 - P_s)$. Thus,

$$\begin{aligned} Z &= \min_{0<s} \left[(1-P_s)e^{-s} + P_s e^{s}\right] \\ &= \left[4P_s(1-P_s)\right]^{1/2} \end{aligned} \tag{A-108}$$

for hard-decision decoding. Substituting this equation into (A-107) with $\alpha = 1$, we obtain

$$P_2(l) \leq \left[4P_s(1-P_s)\right]^{l/2} \tag{A-109}$$

This upper bound has great generality since no specific assumptions have been made about the modulation or coding.

A-2 Convolutional Codes and Trellis Codes

A *convolutional encoder* converts an input of k information bits into an output of n code bits that are Boolean functions of both the current input and preceding information bits. After k bits are shifted into a shift register and k bits are shifted out, n code bits are read out. Each code bit is a Boolean function of the outputs of selected shift-register stages. A convolutional code is *linear* if each Boolean function is a modulo-2 sum because the superposition property applies to the input-output relations and the all-zero codeword is a member of the code. For a linear convolutional code, the minimum Hamming distance between codewords is equal to the minimum Hamming weight of a codeword. The *constraint length* K of a convolutional code is the maximum number of sets of n output bits that can be affected by an input bit. A convolutional encoder is *systematic* if the information bits appear unaltered in each codeword. In contrast to a block codeword, a convolutional codeword can represent an entire message of indefinite length.

A simple example of a nonsystematic linear convolutional encoder with $k = 1$, $n = 2$, and $K = 3$ is shown in Figure A-6(a). The shift register consists of 3 stages, each of which is implemented as a bistable memory element. Information bits enter the shift register in response to clock pulses. After each clock pulse, the most recent information bit becomes the content and output of the first stage, the previous contents of the first two stages are shifted to the right, and the previous content of the third stage is shifted out of the register. The outputs of the modulo-2 adders (exclusive-OR gates) provide two code bits. The *generators* of the output bits are the functions $\mathbf{g}_1 = [1\,0\,1]$ and $\mathbf{g}_2 = [1\,1\,1]$, which indicate the stages that are connected to the adders. In octal form, the two generators are represented by (5, 7). The encoder of a nonsystematic convolutional code with $k = 2$, $n = 3$, and $K = 2$ is shown in Figure A-6(b). In octal form, its generators are (13, 12, 11).

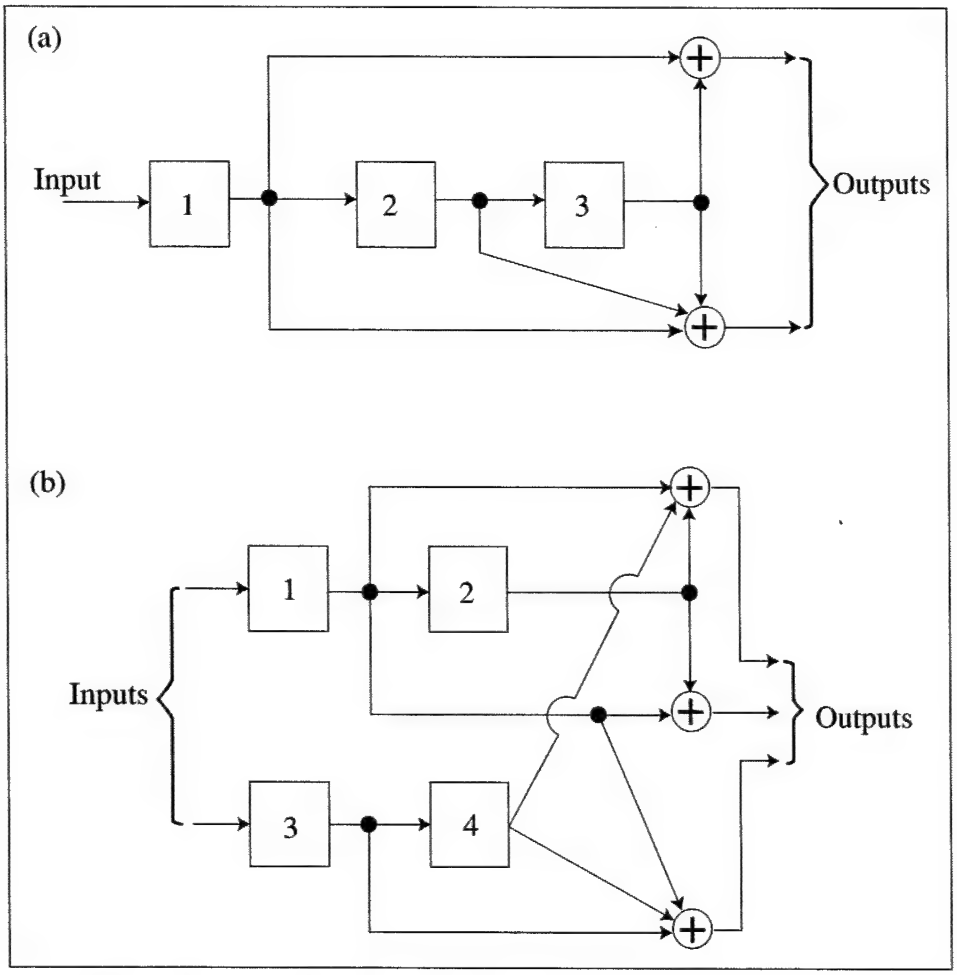


Figure A-6. Encoders of nonsystematic convolutional codes with (a) $K = 3$ and rate = 1/2 and (b) $K = 2$ and rate = 2/3.

Since k bits exit from the shift register as k new bits enter it, only the contents of the first $(K-1)k$ stages prior to the arrival of new bits affect the subsequent output bits of a convolutional encoder. Therefore, the contents of these $(K-1)k$ stages define the *state* of the encoder. The initial state of the encoder is generally the all-zero state. After the message sequence has been encoded $(K-1)k$ zeros must be inserted into the encoder to complete and terminate the codeword. If the number of message bits is much greater than $(K-1)k$, these terminal zeros have a negligible effect and the *code rate* is well approximated by $r = k/n$.

A *trellis diagram* depicts the structure of a convolutional code. A trellis diagram corresponding to the encoder of Figure A-6(a) is shown in Figure A-7. Each of the nodes in a column of a trellis diagram represents the state of the encoder at a specific time prior to a clock pulse. The first bit of a state represents the content of stage 1, while the second bit represents the content of stage 2. Branches connecting nodes represent possible changes of state. Each branch is labeled with the output bits or symbols produced following a clock pulse and the formation of a new encoder state. In this example, the first bit of a branch label refers to the

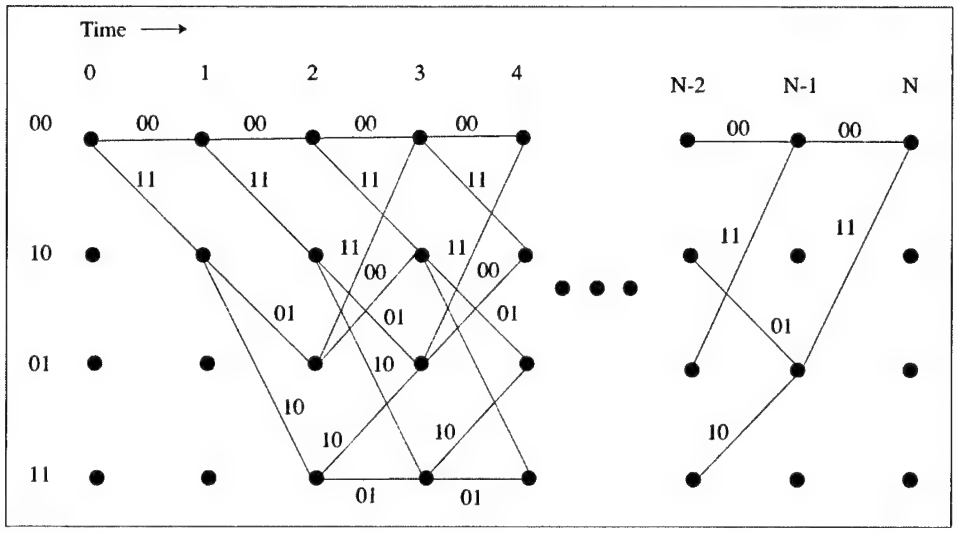


Figure A-7. Trellis diagram corresponding to encoder of Figure A-6.

upper output of the encoder. The upper branch leaving a node corresponds to a 0 input bit, while the lower branch corresponds to a 1.

If the encoder begins in the all-zero state, not all of the other states can be reached until the initial contents have been shifted out. The trellis diagram then becomes identical from column to column until the final $(K-1)k$ input bits force the encoder back to the zero state. Every path from left to right through the trellis represents a possible codeword.

Each branch of the trellis is associated with a branch metric, and the metric of a codeword is defined as the sum of the branch metrics for the path associated with the codeword. A maximum-likelihood decoder selects the codeword with the largest metric (or smallest metric, depending on how branch metrics are defined). The *Viterbi decoder* implements maximum-likelihood decoding efficiently by sequentially eliminating many of the possible paths. At any node, only the partial path reaching that node with the largest partial metric is retained, for any partial path stemming from the node will add the same branch metrics to all paths that merge at that node.

Since the decoding complexity grows exponentially with constraint length, Viterbi decoders are limited to use with convolutional codes of short constraint lengths. A Viterbi decoder for a rate-1/2, $K = 7$ convolutional code has approximately the same complexity as a Reed-Solomon (31,15) decoder. If the constraint length is increased to $K = 9$, the complexity of the Viterbi decoder increases by a factor of approximately 4.

Sequential decoding of convolutional codes [37] is a suboptimal method that does not invariably provide maximum-likelihood decisions. However, because its implementation complexity only weakly depends on the constraint length, very low error probabilities can be attained by using long constraint lengths. The number of computations needed to decode a frame of data is fixed when Viterbi

decoding is used, but is a random variable when sequential decoding is used. When strong interference is present, the excessive computational demands and consequent memory overflows of sequential decoding usually result in a higher P_b than for Viterbi decoding and a much longer decoding delay. Thus, Viterbi decoding is preferable for most communication systems and is assumed in the subsequent analysis of the performance of convolutional codes.

To bound the information-bit error probability for the Viterbi decoder, we assume that the convolutional code is linear and that binary symbols are transmitted. With these assumptions, the distribution of either Hamming or Euclidean distances is invariant to the choice of a reference sequence. Consequently, whether the demodulator makes hard or soft decisions, the assumption that the all-zero sequence is transmitted entails no loss of generality in the derivation of the error probability. Let $a(l, i)$ denote the number of paths diverging at a node from the the correct path, each having Hamming weight l and i incorrect information symbols over the unmerged segment of the path before it merges with the correct path. Thus, the unmerged segment is at Hamming distance l from the correct all-zero segment. Let d_f denote the *minimum free distance*, which is the minimum distance between any two codewords. Although the encoder follows the all-zero path through the trellis, the decoder in the receiver essentially observes successive columns in the trellis, eliminating paths and thereby sometimes introducing errors at each node. The decoder may select an incorrect path that diverges at node j and introduces errors over its unmerged segment. Let $E[N_e(j)]$ denote the expected value of the number of errors introduced at node j. It is known from (A-16) that the information-bit error probability equals the *information-bit error rate*, which is defined as the ratio of the expected number of information-bit errors to the number of information bits applied to the convolutional encoder. Therefore, if there are N branches in a complete path,

$$P_b = \frac{1}{kN} \sum_{j=1}^{N} E[N_e(j)] \tag{A-110}$$

Let $B_j(l, i)$ denote the event that the path with the largest metric diverges at node j and has Hamming weight l and i incorrect information bits over its unmerged segment. Then,

$$E\left[N_e(j)\right] = \sum_{i=1}^{I_j} \sum_{l=d_f}^{D_j} E\left[N_e(j)|B_j(l,i)\right] P\left[B_j(l,i)\right] \tag{A-111}$$

when $E[N_e(j)|B_j(l,i)]$ is the conditional expectation of $N_e(j)$ given event $B_j(l,i)$, $P[B_j(l,i)]$ is the probability of this event, and I_j and D_j are the maximum values of i and l, respectively, that are consistent with the position of node j in the trellis. When $B_j(l,i)$ occurs, i bit errors are introduced into the decoded bits; thus,

$$E[N_e(j)|B_j(l,i)] = i \tag{A-112}$$

Since the decoder may already have departed from the correct path before node j, the union bound gives

$$P\left[B_j(l,i)\right] \leq a(l,i)P_2(l) \tag{A-113}$$

where $P_2(l)$ is the probability that the correct path segment has a smaller metric than an unmerged path segment that differs in l code symbols. Substituting (A-111) to (A-113) into (A-110) and extending the two summations to ∞, we obtain

$$P_b \leq \frac{1}{k}\sum_{i=1}^{\infty}\sum_{l=d_f}^{\infty} ia(l,i)P_2(l) \tag{A-114}$$

The *information-weight spectrum* is defined as

$$B(l) = \sum_{i=1}^{\infty} ia(l,i), \quad l \geq d_f \tag{A-115}$$

In terms of $B(l)$, (A-114) becomes

$$P_b \leq \frac{1}{k}\sum_{l=d_f}^{\infty} B(l)P_2(l) \tag{A-116}$$

For coherent PSK signals over an AWGN channel and soft decisions, (A-45) indicates that

$$P_2(l) = Q\left(\sqrt{\frac{2lr\mathcal{E}_b}{N_0}}\right) \tag{A-117}$$

When the demodulator makes hard decisions and a correct path segment is compared with an incorrect one, correct decoding results if the number of symbol errors in the demodulator output is less than half the number of symbols in which the two segments differ. If the number of symbol errors is exactly half the number of differing symbols, then either of the two segments is chosen with equal probability. Assuming the independence of symbol errors, it follows that for hard-decision decoding

$$P_2(l) = \begin{cases} \displaystyle\sum_{i=(l+1)/2}^{l} \binom{l}{i} P_s^i(1-P_s)^{l-i}, & l \text{ is odd} \\ \displaystyle\sum_{i=l/2+1}^{l} \binom{l}{i} P_s^i(1-P_s)^{l-i} + \frac{1}{2}\binom{l}{l/2}\left[P_s(1-P_s)\right]^{l/2}, & l \text{ is even} \end{cases} \tag{A-118}$$

Soft-decision decoding typically provides a 2 dB power savings at $P_b = 10^{-5}$ compared to hard-decision decoding for communications over the AWGN

channel. Since the loss due to even three-bit quantization usually is 0.2 to 0.3 dB, soft-decision decoding is highly preferable.

Among the convolutional codes of a given code rate and constraint length, the one giving the smallest upper bound in (A-116) can sometimes be determined by a complete computer search. The codes with the largest value of d_f are selected, and the *catastrophic codes*, for which a finite number of demodulated symbol errors can cause an infinite number of decoded information-bit errors, are eliminated. All remaining codes that do not have the minimum value of $B(d_f)$ are eliminated. If more than one code remains, codes are eliminated on the basis of the minimal values of $B(d_f + 1), B(d_f + 2), \ldots$, until one code remains. For binary codes of rates 1/2, 1/3, and 1/4, codes with these favorable distance properties have been determined [41]. For these codes and constraint lengths up to 12, Tables A-3, A-4, and A-5 list the corresponding values of d_f and $B(d_f + i), i = 0, 1, \ldots, 7$. Also listed in octal form are the generator sequences that determine which shift-register stages feed the modulo-two adders a associated with each code bit. For example, the best $K = 3$, rate-1/2 code in Table A-3 has generators 5, 7, which specify the connections illustrated in Figure A-6(a).

Table A-3. Parameter Values for Convolutional Codes with Rate = 1/2 and Favorable Distance Properties.

			$B(d_f + i)$ for $i = 0, 1, \ldots, 6$						
K	d_f	Generators	0	1	2	3	4	5	6
3	5	5, 7	1	4	12	32	80	192	448
4	6	15, 17	2	7	18	49	130	333	836
5	7	23, 35	4	12	20	72	225	500	1324
6	8	53, 75	2	36	32	62	332	701	2342
7	10	133, 171	36	0	211	0	1404	0	11,633
8	10	247, 371	2	22	60	148	340	1008	2642
9	12	561, 763	33	0	281	0	2179	0	15,035
10	12	1131, 1537	2	21	100	186	474	1419	3542
11	14	2473, 3217	56	0	656	0	3708	0	27,518
12	15	4325, 6747	66	98	220	788	2083	5424	13,771

Table A-4. Parameter Values for Convolutional Codes with Rate = 1/3 and Favorable Distance Properties.

			$B(d_f + i)$ for $i = 0, 1, \ldots, 6$						
K	d_f	Generators	0	1	2	3	4	5	6
3	8	5, 7, 7	3	0	15	0	58	0	201
4	10	13, 15, 17	6	0	6	0	58	0	118
5	12	25, 33, 37	12	0	12	0	56	0	320
6	13	47, 53, 75	1	8	26	20	19	62	86
7	15	117, 127, 155	7	8	22	44	22	94	219
8	16	225, 331, 367	1	0	24	0	113	0	287
9	18	575, 673, 727	2	10	50	37	92	92	274
10	20	1167, 1375, 1545	6	16	72	68	170	162	340
11	22	2325, 2731, 3747	17	0	122	0	345	0	1102
12	24	5745, 6471, 7553	43	0	162	0	507	0	1420

Table A-5. Parameter Values for Convolutional Codes with Rate = 1/4 and Favorable Distance Properties.

			$B(d_f + i)$ for $i = 0, 1, \ldots, 6$						
K	d_f	Generators	0	1	2	3	4	5	6
3	10	5, 5, 7, 7	1	0	4	0	12	0	32
4	13	13, 13, 15, 17	4	2	0	10	3	16	34
5	16	25, 27, 33, 37	8	0	7	0	17	0	60
6	18	45, 53, 67, 77	5	0	19	0	14	0	70
7	20	117, 127, 155, 171	3	0	17	0	32	0	66
8	22	257, 311, 337, 355	2	4	4	24	22	33	44
9	24	533, 575, 647, 711	1	0	15	0	56	0	69
10	27	1173, 1325, 1467, 1751	7	10	0	28	54	58	54

Approximate upper bounds on P_b for rate-1/2, rate-1/3, and rate-1/4 convolutional codes with coherent PSK, soft-decision decoding, and infinitely fine quantization are depicted in Figures A-8 to A-10. The curves are computed by using (A-117), $k = 1$, and Tables A-3 to A-5 in (A-116) and then truncating the series after seven terms. This truncation gives a tight upper bound in P_b for $P_b \leq 10^{-2}$. However, the truncation may exclude significant contributions to the upper bound when $P_b > 10^{-2}$, and the bound itself becomes looser as P_b increases. The figures indicate that the code performance improves with increases in the constraint length and as the code rate decreases if $K \geq 4$. The decoder complexity is almost exclusively dependent on K because there are 2^{K-1} encoder states. However, as the code rate decreases, more bandwidth and a more difficult bit synchronization are required.

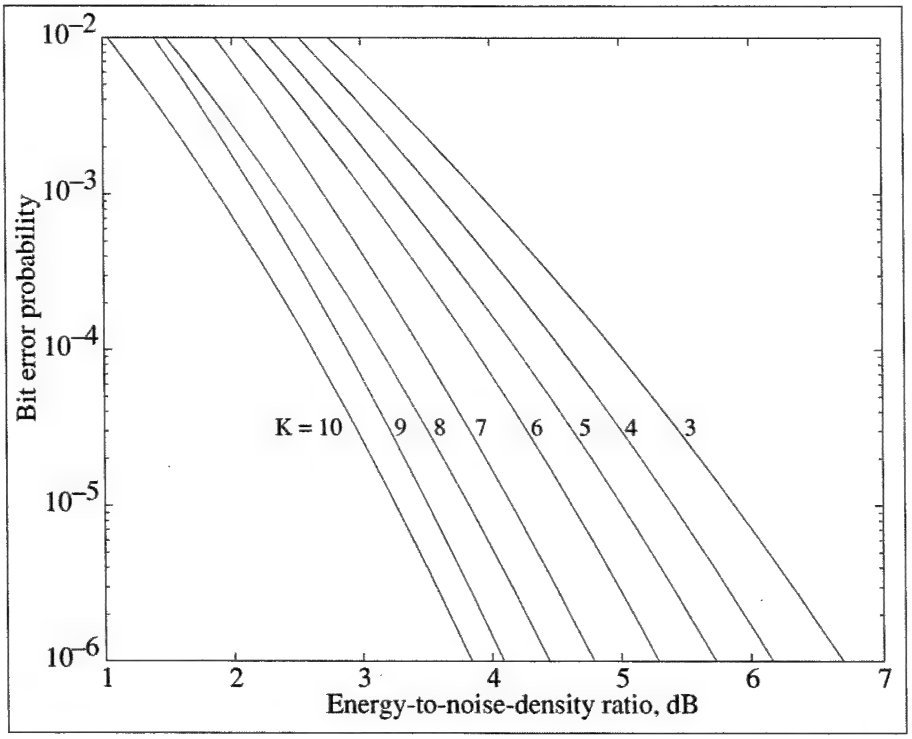


Figure A-8. Information-bit error probability for rate = 1/2 convolutional codes with different constraint lengths and coherent PSK.

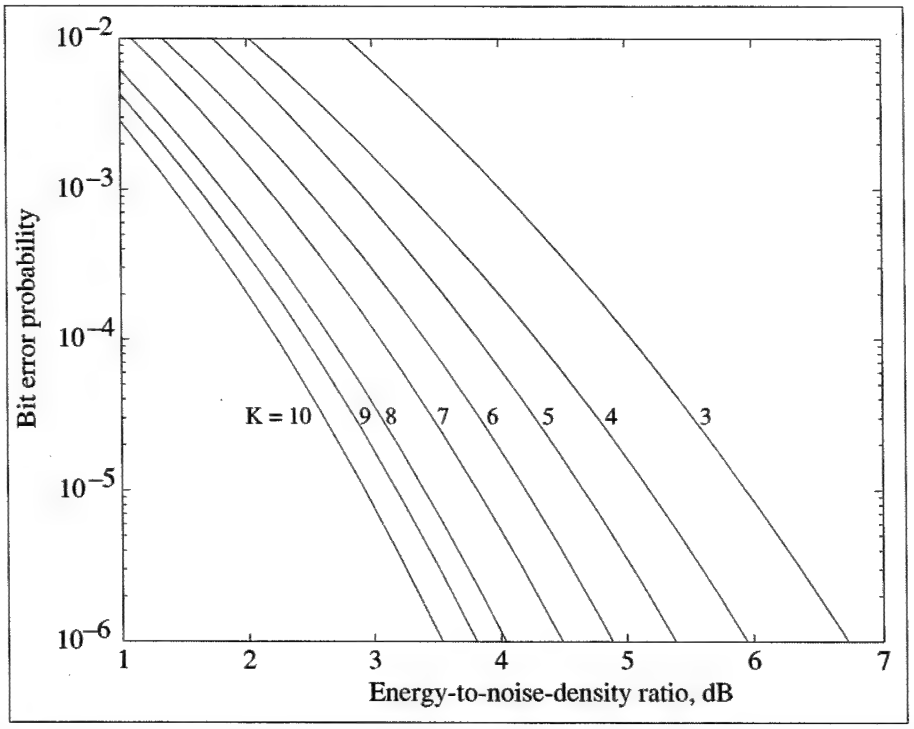


Figure A-9. Information-bit error probability for rate = 1/3 convolutional codes with different constraint lengths and coherent PSK.

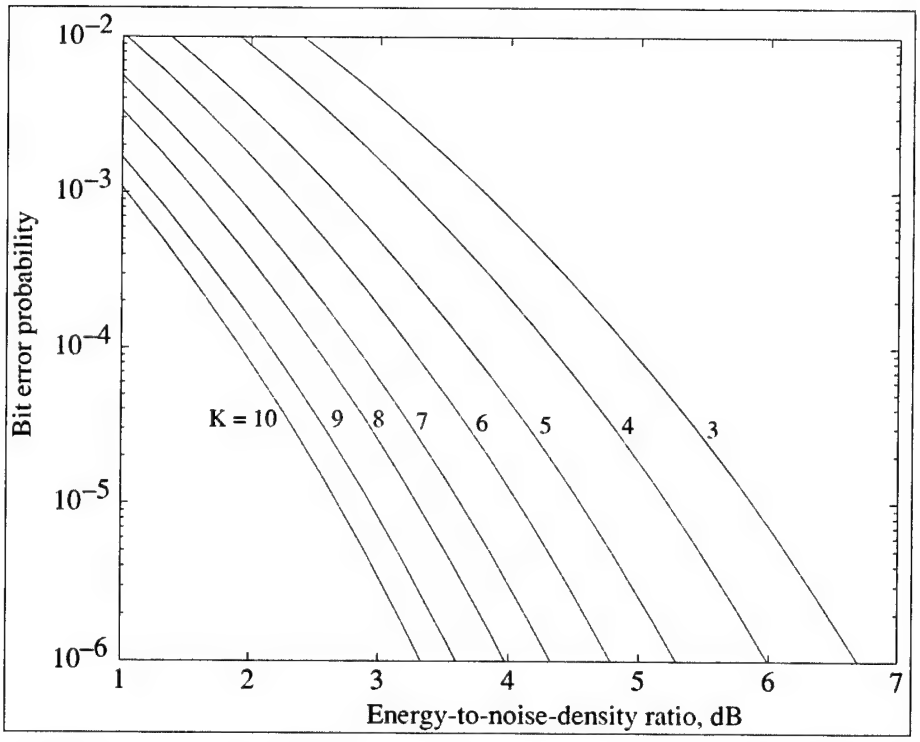


Figure A-10. Information-bit error probability for rate = 1/4 convolutional codes with different constraint lengths and coherent PSK.

For convolutional codes of rate $1/n$, two trellis branches enter each state. For higher-rate codes with k information bits per branch, 2^k trellis branches enter each state and the computational complexity may be large. This complexity can be avoided by using *punctured convolutional codes*. These codes are generated by periodically deleting bits from one or more output streams of an encoder for an unpunctured rate-$1/n$ code. For a period-p punctured code, p sets of n bits are written into a buffer from which $p + \nu$ bits are read out, where $1 \leq \nu \leq (n-1)p$. Thus, a punctured convolutional code has a rate of the form

$$r = \frac{p}{p+\nu} \quad , \quad 1 \leq \nu \leq (n-1)p \tag{A-119}$$

The decoder of a punctured code uses the same decoder and trellis as the parent code, but uses only the metrics of the unpunctured bits as it proceeds through the trellis. The upper bound on P_b is given by (A-116) with $k = 1$. For most code rates, there are punctured codes with the largest minimum free distance of any convolutional code with that code rate. Punctured convolutional codes enable the efficient implementation of a variable-rate error-control system with a single encoder and decoder. However, the periodic character of the trellis of a punctured code requires that the decoder acquire frame synchronization.

Coded nonbinary sequences can be produced by converting the outputs of a binary convolutional encoder into a single nonbinary symbol, but this procedure does not optimize the nonbinary code's Hamming distance properties. Better nonbinary codes, such as the dual-k codes, are possible [38] but do not provide

as good a performance as the nonbinary Reed-Solomon codes with the same transmission bandwidth.

In principle, $B(l)$ can be determined from the *generating function*, $T(D, I)$, which can be derived for some convolutional codes by treating the state diagram as a signal flow graph [36], [37]. The generating function is a polynomial in D and I of the form

$$T(D, I) = \sum_{i=1}^{\infty} \sum_{l=d_f}^{\infty} a(l, i) D^l I^i \tag{A-120}$$

where $a(l, i)$ represents the number of distinct unmerged segments characterized by l and i. The derivative at $I = 1$ is

$$\left.\frac{\partial T(D, I)}{\partial I}\right|_{I=1} = \sum_{i=1}^{\infty} \sum_{l=d_f}^{\infty} i a(l, i) D^l = \sum_{l=d_f}^{\infty} B(l) D^l \tag{A-121}$$

Thus, the bound on P_b given by (A-116), is determined by substituting $P_2(l)$ in place of D^l in the polynomial expansion of the derivative of $T(D, I)$ and multiplying the result by $1/k$. In many applications, it is possible to establish an inequality of the form

$$P_2(l) \leq \alpha Z^l \tag{A-122}$$

where α and Z are independent of l. It then follows from (A-116), (A-121), and (A-122) that

$$P_b \leq \frac{\alpha}{k} \left.\frac{\partial T(D, I)}{\partial I}\right|_{I=1, D=Z} \tag{A-123}$$

For soft-decision decoding and coherent PSK, $P_2(l)$ is given by (A-117). Using the definition of $Q(x)$ given by (A-30), changing variables, and comparing the two sides of the following inequality, we verify that

$$\begin{aligned} Q\left(\sqrt{\nu + \beta}\right) &= \frac{1}{\sqrt{2\pi}} \int_0^{\infty} \exp\left[-\frac{1}{2}\left(y + \sqrt{\nu + \beta}\right)^2\right] dy \\ &\leq \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\beta}{2}\right) \int_0^{\infty} \exp\left[-\frac{1}{2}(y + \sqrt{\nu})^2\right] dy , \\ &\quad \nu \geq 0 , \quad \beta \geq 0 \end{aligned} \tag{A-124}$$

A change of variables yields

$$Q\left(\sqrt{\nu + \beta}\right) \leq \exp\left(-\frac{\beta}{2}\right) Q\left(\sqrt{\nu}\right) , \quad \nu \geq 0, \quad \beta \geq 0 \tag{A-125}$$

Substituting this inequality into (A-117) with the appropriate choices for ν and β gives

$$P_2(l) \leq Q\left(\frac{\sqrt{2 d_f r \mathcal{E}_b}}{N_0}\right) \exp\left[-(l - d_f) r \mathcal{E}_b / N_0\right] \tag{A-126}$$

Thus, the upper bound on $P_2(l)$ may be expressed in the form given by (A-122) with

$$\alpha = Q\left(\frac{\sqrt{2d_f r\mathcal{E}_b}}{N_0}\right)\exp(d_f r\mathcal{E}_b/N_0) \tag{A-127}$$

$$Z = \exp(-r\mathcal{E}_b/N_0) \tag{A-128}$$

For other channels, codes, and modulations, an upper bound on $P_2(l)$ in the form given by (A-122) can often be derived from the Chernoff bound.

Trellis-Coded Modulation

To add an error-correcting code to a communication system, while avoiding a bandwidth expansion, one may increase the number of signal constellation points. For example, if a rate-2/3 code is added to a system using QPSK, then the bandwidth is preserved if the modulation is changed to eight-phase PSK (8-PSK). Since each symbol of the latter modulation represents 3/2 as many bits as a QPSK symbol, the channel-symbol rate is unchanged. The problem is that the change from QPSK to the more compact 8-PSK constellation causes an increase in the channel-symbol error probability that cancels most of the decrease due to the encoding. To overcome this problem, the coding and modulation processes must be integrated.

Trellis-coded modulation, the most widely used modulation that incorporates coding, is usually produced by a system of the form shown in Figure A-11. For $k > 1$, each input of k information bits is divided into two groups. One group of k_1 bits is applied to a convolutional encoder while the other group of $k_2 = k - k_1$ bits remains uncoded. The $k_1 + 1$ output bits of the convolutional encoder select one of 2^{k_1+1} possible subsets of the points in the constellation of the modulator. The k_2 uncoded bits select one of 2^{k_2} points in the chosen subset. If $k_2 = 0$, there are no uncoded bits and the convolutional encoder output bits select the constellation point. Each constellation point is a complex number representing an amplitude and phase.

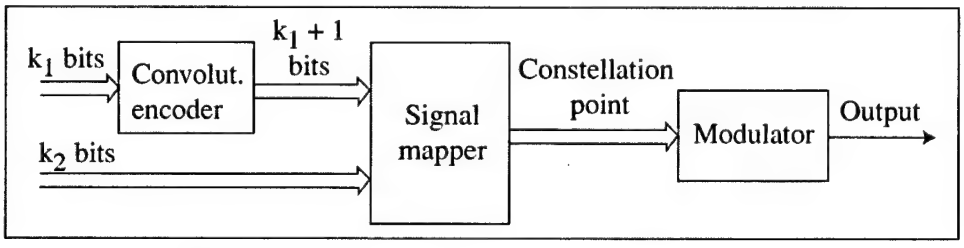


Figure A-11. Encoder for trellis-coded modulation.

For example, suppose that $k = 2$, $k_1 = k_2 = 1$, and $n = 2$ in the encoder of Figure A-11, and an 8-PSK modulator produces an output from a constellation of 8 points. Each of the four subsets that may be selected by the two convolutional-code bits comprises two antipodal points in the 8-PSK constellation, as shown in

Figure A-12. If the convolutional encoder has the form of Figure A-6(a), then the trellis of Figure A-7 illustrates the state transitions of both the underlying convolutional code and the trellis code. The presence of the single uncoded bit implies that each transition between states in the trellis corresponds to two different transitions and two different phases of the transmitted 8-PSK waveform.

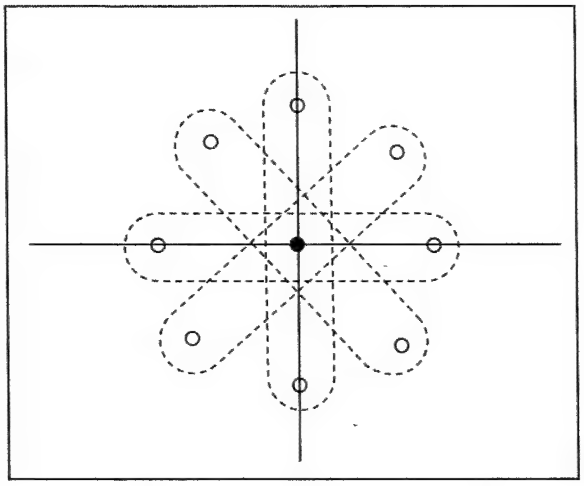


Figure A-12. The constellation of 8-PSK symbols partitioned into 4 subsets.

In general, there are 2^{k_2} parallel transitions between every pair of states in the trellis. Often, the dominant error events consist of mistaking one of these parallel transitions for the correct one. If the symbols corresponding to parallel transitions are separated by large Euclidean distances and the constellation subsets associated with transitions are suitably chosen, then the trellis-coded modulation with soft-decision Viterbi decoding can yield a substantial performance improvement [36], [37], [38]. This improvement usually ranges from 4 to 6 dB, depending on the number of states and, hence, the implementation complexity. The minimum Euclidean distance between a correct trellis-code path and an incorrect one is called the *free Euclidean distance* and is denoted by $d_{fe}\sqrt{\mathcal{E}_s}$. Let B_{fe} denote the total number of information bit errors associated with erroneous paths that are at the free Euclidian distance from the correct path. The latter paths dominate the error events when the SNR is high. An analysis similar to the one for convolutional codes indicates that for the AWGN channel and a high SNR,

$$P_b \approx \frac{B_{fe}}{k} Q\left(\sqrt{\frac{d_{fe}^2 r \mathcal{E}_b}{2N_0}}\right) \tag{A-129}$$

A-3 Interleaving

An *interleaver* is a device that permutes the order of a sequence of symbols. A *deinterleaver* is the corresponding device that restores the original order of the sequence. A major application is the interleaving of modulated symbols

transmitted over a communication channel. After deinterleaving at the receiver, a burst of channel symbol errors or corrupted symbols is dispersed over a number of codewords or constraint lengths, thereby facilitating the removal of the errors by the decoding. Ideally, the interleaving and deinterleaving ensures that the decoder encounters statistically independent symbol decisions or metrics, as it would if the channel were memoryless. Interleaving of channel symbols is useful when error bursts are caused by fast fading, interference, or even decision-directed equalization.

A *block interleaver* performs identical permutations on successive blocks of symbols. As illustrated in Figure A-13, mn successive input symbols are stored in a random-access memory (RAM) as a matrix of m rows and n columns. The input sequence is written into the interleaver in successive rows, but successive columns are read to produce the interleaved sequence. Thus, if the input sequence is numbered $1, 2, \ldots, n, n+1, \ldots, mn$, the interleaved sequence is $1, n+1, 2n+1, \ldots, 2, n+2, \ldots, mn$. For continuous interleaving, two RAMs are needed. Symbols are written into one RAM matrix while previous symbols are read from the other. In the deinterleaver, symbols are stored by column in one matrix, while previous symbols are read by rows from another. Consequently, a delay of $2mnT_s$ must be accommodated and synchronization is required at the deinterleaver.

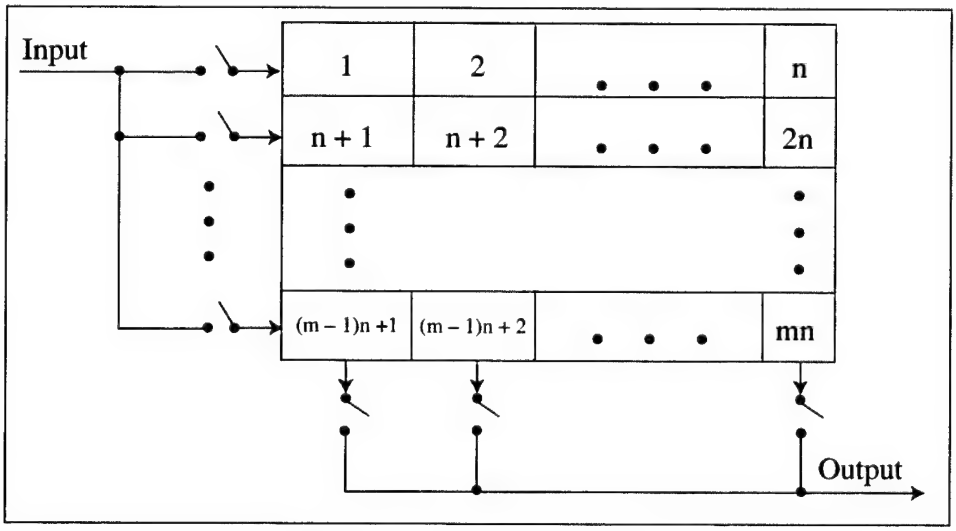


Figure A-13. Block interleaver.

When channel symbols are interleaved, the parameter n equals or exceeds the block codeword length or a few constraint lengths of a convolutional code. Consequently, if a burst of m or fewer consecutive symbol errors occurs and there are no other errors, then each block codeword or constraint length, after deinterleaving, has at most one error, which can be eliminated by the error-correcting code. Similarly, a block code that can correct t errors is capable of correcting a single burst of errors spanning as many as mt symbols. Since fading

can cause correlated errors, it is necessary that mT_s exceed the channel coherence time. Interleaving effectiveness can be thwarted by slow fading that cannot be accommodated without large buffers that cause an unacceptable delay.

Other types of interleavers that are closely related to the block interleaver include the *convolutional interleaver* and the *helical interleaver* [15]. A helical interleaver reads symbols from its matrix diagonally instead of by column in such a way that consecutive interleaved symbols are never read from the same row or column. Both helical and convolutional interleavers and their corresponding deinterleavers confer advantages in certain applications, but do not possess the inherent simplicity and compatibility with block structures that block interleavers have.

A *pseudorandom interleaver* permutes each block of symbols pseudorandomly. Pseudorandom interleavers may be applied to channel symbols, but their main application is as critical elements in turbo encoders and encoders of serial concatenated codes that use iterative decoding (Section A-4).The desired permutation may be stored in a read-only memory (ROM) as a sequence of addresses or permutation indices. Each block of symbols is written sequentially into a RAM matrix and then interleaved by reading them in the order dictated by the contents of the ROM.

If the interleaver is large, it is often preferable to generate the permutation indices by an algorithm rather than storing them in a ROM. If the interleaver size is $N = mn = 2^\nu - 1$, then a linear feedback shift register with ν stages that produces a maximal-length sequence can be used. The binary outputs of the shift-register stages constitute the *state* of the register. The state specifies the index from 1 to N that defines a specific interleaved symbol. The shift register generates all N states and indices periodically.

An *S-random interleaver* is a pseudorandom interleaver that constrains the minimum interleaving distance. A tentative permutation index is compared with the S previously selected indices, where $1 \leq S < N$. If the tentative index does not differ in absolute value from the S previous ones by at least S, then it is discarded and replaced by a new tentative index. If it does, then the tentative index becomes the next selected index. This procedure continues until all N pseudorandom indices are selected. The S-random interleaver is frequently used in turbo or serial concatenated encoders.

A-4 Concatenated Codes

A *concatenated code* uses multiple levels of coding to achieve a large error-correcting capability with manageable implementation complexity by breaking the decoding process into stages. In practice, two levels of coding have been found to be effective. Figure A-14 is a functional block diagram of a communication system incorporating a concatenated code. The channel interleaver permutes the code bits to ensure the random distribution of code-bit

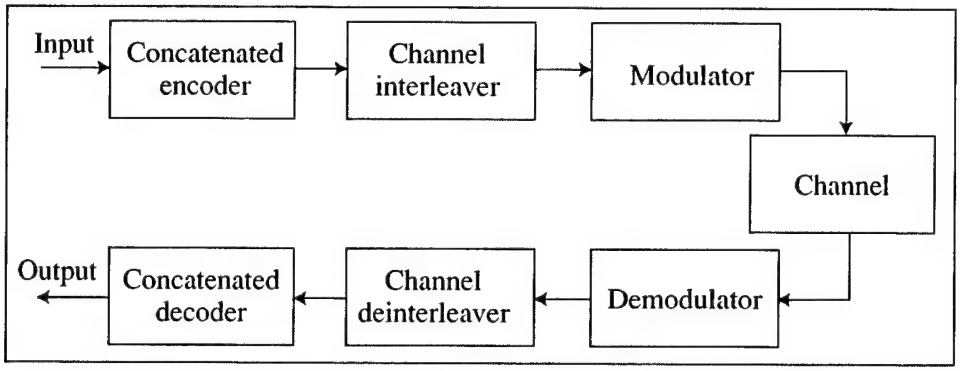


Figure A-14. Concatenated coding in transmitter and receiver.

errors at the input of the concatenated decoder. Concatenated codes may be classified as classical concatenated codes, turbo codes, or serial concatenated codes.

Classical Concatenated Codes

Classical concatenated codes are serial concatenated codes with the encoder and decoder forms shown in Figure A-15. In the most common configuration for classical concatenated codes, an *inner code* uses binary symbols and a Reed-Solomon *outer code* uses nonbinary symbols. The outer-encoder output symbols are interleaved, and then these nonbinary symbols are converted into binary symbols that are encoded by the inner encoder. In the receiver, a grouping of the binary inner-decoder output symbols into nonbinary outer-code symbols is followed by symbol deinterleaving that disperses the outer-code symbol errors. Consequently, the outer decoder is able to correct most symbol errors originating in the inner-decoder output. The concatenated code has rate

$$r = r_1 r_0 \tag{A-130}$$

where r_1 is the inner-code rate and r_0 is the outer-code rate.

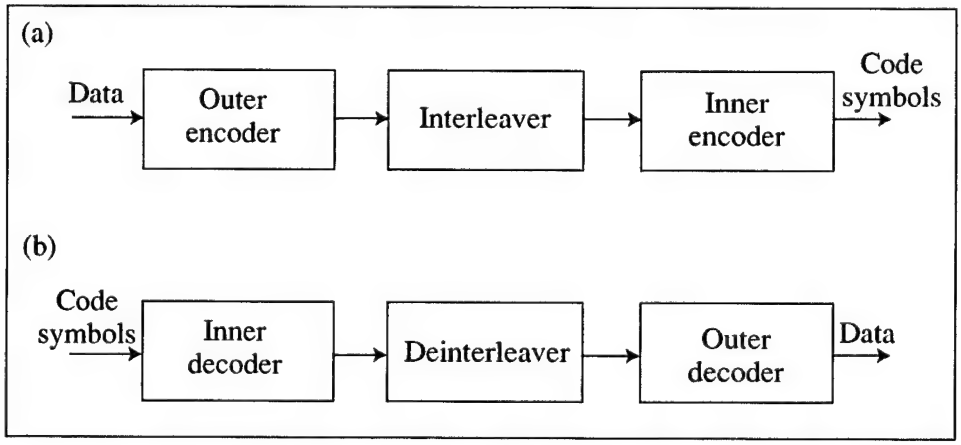


Figure A-15. Structure of serial concatenated code: (a) encoder and (b) classical decoder.

A variety of inner codes have been proposed. The dominant and most powerful concatenated code of this type comprises a binary convolutional inner code and a Reed-Solomon outer code. At the output of a convolutional inner decoder using the Viterbi algorithm, the bit errors occur over spans with an average length that depends on the signal-to-noise ratio. The deinterleaver is designed to ensure that Reed-Solomon symbols formed from bits in the same typical error span do not belong to the same Reed-Solomon codeword. In the worst case, the inner decoder produces bit errors that are separated enough that each one causes a separate symbol error at the input to the Reed-Solomon decoder. Since there are m times as many bits as symbols, the symbol error probability P_{s1} is upper bounded by m times the bit error probability at the inner-decoder output, where $m = \log_2 q$ is the number of bits in a Reed-Solomon code symbol, and is lower bounded by this bit error probability. Thus, for binary convolutional inner codes,

$$\frac{1}{k}\sum_{l=d_f}^{\infty} B(l)P_2(l) \le P_{s1} \le \frac{\log_2 q}{k}\sum_{l=d_f}^{\infty} B(l)P_2(l) \qquad \text{(A-131)}$$

where $P_2(l)$ is given by (A-107) and (A-106). Assuming that the deinterleaving ensures independent symbol errors at the outer-decoder input, and that the Reed-Solomon code is loosely packed, (A-26) and (A-27) imply that

$$P_b \approx \frac{q}{2(q-1)}\sum_{i=t+1}^{n}\binom{n-1}{i-1}P_{s1}^{i}(1-P_{s1})^{n-1} \qquad \text{(A-132)}$$

For coherent PSK modulation with soft decisions, $P_2(l)$ is given by (A-116); if hard decisions are made, (A-118) applies.

Figure A-16 depicts examples of the approximate upper bound on the performance in white Gaussian noise of concatenated codes with coherent PSK, soft demodulator decisions, an inner binary convolutional code with k = 1, K = 7, and rate = 1/2, and various Reed-Solomon outer codes. Equation (A-132) and the upper bound in (A-131) are used. The bandwidth required by a concatenated code is B/r, where B is the uncoded PSK bandwidth. Since (A-130) gives $r < 1/3$, the codes of the figure require less bandwidth than rate-1/3 convolutional codes.

Turbo Codes

Turbo codes are parallel concatenated codes that use iterative decoding [36], [42], [43]. As shown in Figure A-17, the encoder of a turbo code has two component encoders, one of which directly encodes the information bits while the other encodes interleaved bits. The iterative decoding requires that both component codes be systematic and of the same type, that is, both convolutional or both block.

A *turbo convolutional code* uses two binary convolutional codes as its component codes. The multiplexer output comprises both the information and

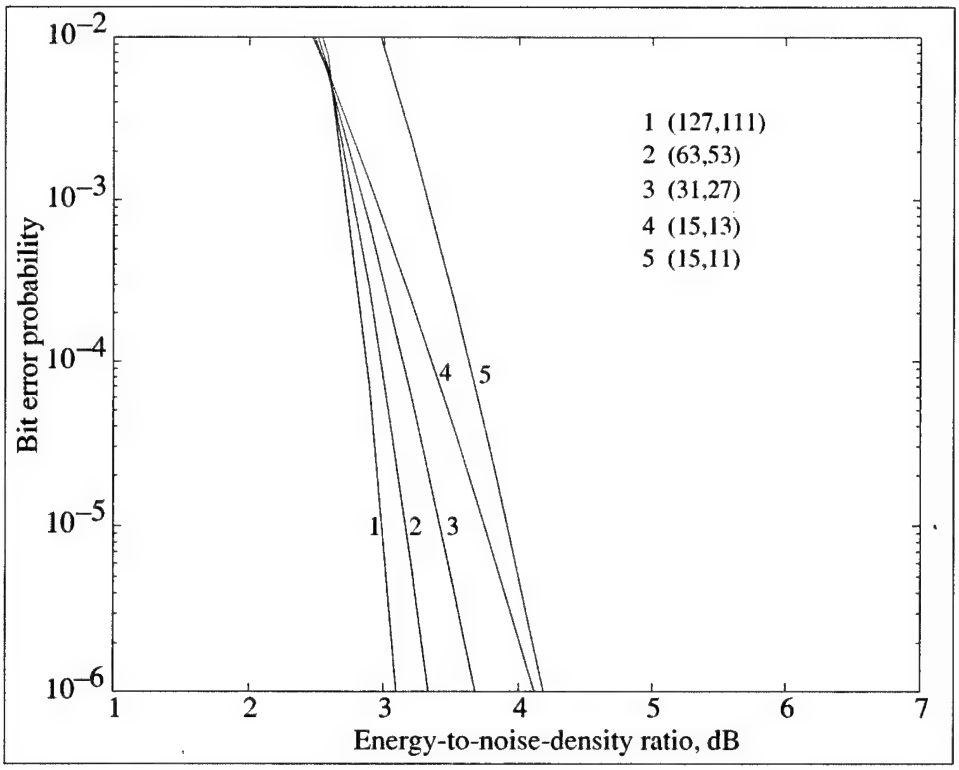


Figure A-16. Information-bit error probability for concatenated codes with inner convolutional code ($K = 7$, rate = 1/2), various Reed-Solomon (n, k) outer codes, and coherent PSK.

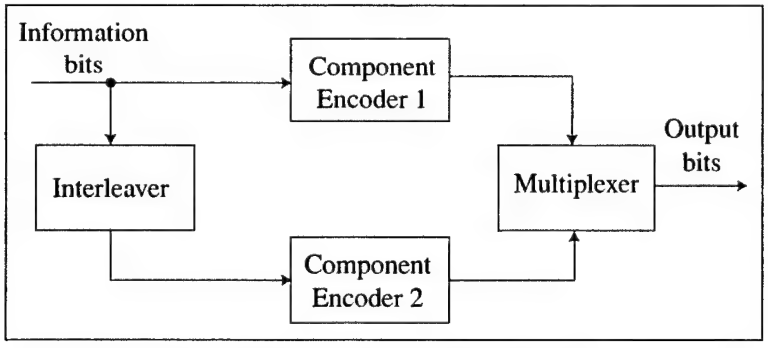


Figure A-17. Encoder of turbo code.

parity bits produced by encoder 1 but only the parity bits produced by encoder 2. Because of their superior distance properties, *recursive systematic convolutional encoders* are used in turbo encoders [36]. Each of these encoders has feedback that causes the shift-register state to depend on its previous outputs. Usually, identical rate-1/2 component codes are used, and a rate-1/3 turbo code is produced. However, if the multiplexer punctures the parity streams, a higher rate of 1/2 or 2/3 can be obtained. Although it requires frame synchronization in the decoder, the puncturing may serve as a convenient means of adapting the code rate to the channel conditions. The purpose of the interleaver, which may be a block or pseudorandom interleaver, is to permute the input bits of encoder 2 so

that it is unlikely that both component codewords will have a low weight. Thus, the turbo code has very few low-weight codewords, whether or not its free distance is large. Terminating tail bits are inserted into both component convolutional codes so that the turbo trellis terminates in the all-zero state and the turbo code can be treated as a block code.

To produce a rate-1/2 turbo code from rate-1/2 convolutional component codes, alternate puncturing of the even parity bits of encoder 1 and the odd parity bits of encoder 2 is done. Consequently, an odd information bit has its associated parity bit of code 1 transmitted. However, because of the interleaving that precedes encoder 2, an even information bit may have neither its associated parity bit of code 1 nor that of code 2 transmitted. Instead, some odd information bits may have both associated parity bits transmitted, although not successively because of the interleaving. Since some information bits have no associated parity bits transmitted, the decoder is less likely to be able to correct errors in those information bits. A convenient means of avoiding this problem, and ensuring that exactly one associated parity bit is transmitted for each information bit, is to use a block interleaver with an odd number of rows and an odd number of columns. This procedure, or any other that separates the odd and even information bits, is called *odd-even separation*. Simulation results confirm that odd-even separation improves the system performance when puncturing and block interleavers are used, but odd-even separation is not beneficial in the absence of puncturing [43]. In a system with a small interleaver size, block interleavers with odd-even separation usually give a better system performance than pseudorandom interleavers, but the latter are usually superior when the interleaver size is large.

The interleaver size is equal to the block length or frame length of the codes. As the block length increases, the performance of a turbo convolutional code often approaches within less than 1 dB of the information-theoretic limit. However, as the block length increases, so does the *system latency*, which is the delay between the input and final output. The potentially large system latency and the system complexity are the primary disadvantages of turbo codes.

A maximum-likelihood decoder minimizes the probability that a received codeword or an entire received sequence is in error. A turbo decoder is designed to minimize the error probability of each information bit. Under either criterion, an optimal decoder would use the sampled demodulator output streams for the information bits and the parity bits of both component codes. A turbo decoder comprises separate component decoders for each component code, which is crucial in reducing the decoder complexity. Each component decoder uses a version of the *maximum a posteriori* (MAP) or *BCJR* algorithm proposed by Bahl, Cocke, Jelinek, and Raviv [36], [43]. As shown in Figure A-18, component decoder 1 of a turbo decoder is fed by demodulator outputs corresponding to the information bits and the parity bits of code 1, while decoder 2 is fed by outputs corresponding to the information bits and the parity bits of code 2. For each information bit u_k, the MAP algorithm computes the log-likelihood ratio (LLR)

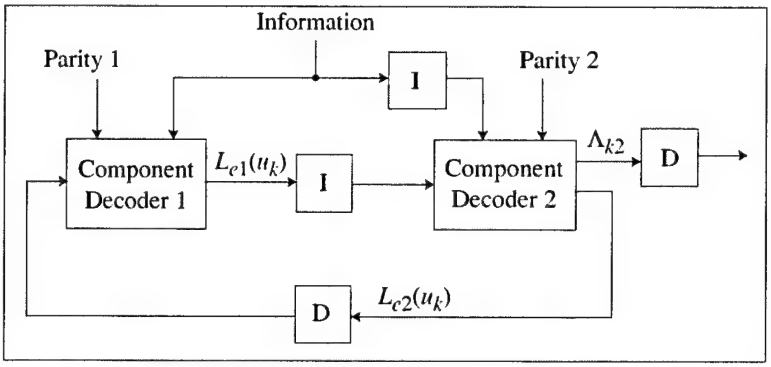


Figure A-18. Decoder of turbo code. I = interleaver; D = deinterleaver.

of the probabilities that this bit is +1 or -1 given the vector **y** of demodulator outputs applied to the component decoder:

$$\Lambda_k = \ln\left[\frac{P(u_k = +1|\mathbf{y})}{P(u_k = -1|\mathbf{y})}\right] \tag{A-133}$$

The LLRs of the information bits are iteratively updated in the two component decoders by passing information between them.

It can be shown that for communications over an AWGN or fading channel using coherent PSK, Λ_k for each component decoder can be decomposed as [43]

$$\Lambda_k = L(u_k) + L(y_{sk}|u_k) + L_e(u_k) \tag{A-134}$$

where the *a priori* LLR is initially

$$L(u_k) = \ln\left[\frac{P(u_k = +1)}{P(u_k = -1)}\right] \tag{A-135}$$

y_{sk} is the demodulator output corresponding to the systematic or information bit u_k, and the *extrinsic information* $L_e(u_k)$ is a function of the demodulated output sequence **y** excluding y_{sk}, but including outputs corresponding to the parity bits processed by the component decoder. The term $L(y_{sk}|u_k)$, which represents information about u_k provided by y_{sk}, is defined as

$$L(y_{sk}|u_k) = \ln\left[\frac{f(y_{sk}|u_k = +1)}{f(y_{sk}|u_k = -1)}\right] \tag{A-136}$$

where $f(y_{sk}|u_k = j)$ is the conditional density of y_{sk} given that $u_k = j$. Let N_{0k} denote the noise power spectral density associated with u_k. For coherent PSK, (A-41) with $y_i \rightarrow y_{sk}$, $N_{0i} \rightarrow N_{0k}$, and $x_{li} \rightarrow \alpha u_k$, where α accounts for the fading attenuation, gives the conditional density

$$f(y_{sk}|u_k = \pm 1) = \frac{1}{\sqrt{\pi N_{0k}/2}} \exp\left[-\frac{(y_{sk} \mp \sqrt{\mathcal{E}_s/2}\alpha)^2}{N_{0k}/2}\right] \tag{A-137}$$

Substitution into (A-136) yields

$$L(y_{sk}|u_k) = L_c y_{sk} \ , \quad L_c = 4\alpha \frac{\sqrt{2\mathcal{E}_s}}{N_{0k}} \tag{A-138}$$

The *channel reliability factor* L_c must be known or estimated to compute Λ_k.

Since almost always no *a priori* knowledge of the likely value of the bit u_k is available, $P(u_k) = 0.5$ is assumed, and $L(u_k)$ is set to zero for the first iteration of component decoder 1. However, for subsequent iterations of either component decoder, $L(u_k)$ for one decoder is set equal to the extrinsic information $L_e(u_k)$ calculated by the other decoder at the end of its previous iteration. As indicated by (A-134), $L_e(u_k)$ can be calculated by subtracting $L(u_k)$ and $L_c y_{sk}$ from Λ_k, which is computed by the MAP algorithm. Since the extrinsic information depends primarily on the constraints imposed by the code used, it provides additional information to the decoder to which it is transfered. As indicated in Figure A-18, appropriate interleaving or deinterleaving is required to ensure that the extrinsic information $L_{e1}(u_k)$ or $L_{e2}(u_k)$ is applied to each component decoder in the correct sequence. When the iterative process terminates, the LLR Λ_{k2} from component decoder 2 is deinterleaved and then applied to a device that makes a hard decision. Performance improves with the number of iterations, but simulation results indicate that little is gained beyond roughly 4 to 12 iterations.

The generic name for a version of the MAP algorithm or an approximation of it is *soft-in soft-out (SISO) algorithm*. The *log-MAP algorithm* is an SISO algorithm that transforms the MAP algorithm into the logarithmic domain, thereby simplifying operations and reducing numerical problems while causing no performance degradation. The *max-log-MAP algorithm* and the *soft-output Viterbi algorithm (SOVA)* are SISO algorithms that reduce the complexity of the log-MAP algorithm at the cost of some performance degradation [36], [43]. The max-log-MAP algorithm is roughly 2/3 as complex as the log-MAP algorithm and typically degrade the performance by 0.1 dB to 0.2 dB at $P_b = 10^{-4}$. The SOVA algorithm is roughly 1/3 as complex as the log-MAP algorithm and typically degrades the performance by 0.5 dB to 1.0 dB at $P_b = 10^{-4}$. The MAP, log MAP, max-log-MAP, and SOVA algorithms have complexities that increase linearly with the number of states of the component codes.

The log-MAP algorithm is roughly 4 times as complex as the standard Viterbi algorithm [43]. For 2 identical component decoders and typically 8 algorithm iterations, the overall complexity of a turbo decoder is roughly 64 times that of a Viterbi decoder for one of the component codes. The complexity of the decoder increases while the performance improves as the constraint length K of each component code increases. The complexity of a turbo decoder using 8 iterations and component convolutional codes with $K = 3$ is approximately the same as that of a Viterbi decoder for a convolutional code with $K = 9$.

If N_{0k} is unknown and may be significantly different from symbol to symbol, a standard procedure [31] is to replace the LLR of (A-136) with the *generalized*

log-likelihood ratio

$$L(y_{sk}|u_k) = \ln\left[\frac{f(y_{sk}|u_k = +1, N_1)}{f(y_{sk}|u_k = -1, N_2)}\right] \tag{A-139}$$

where N_1 and N_2 are maximum-likelihood estimates of N_{0k} in the corresponding densities. From (A-137), we obtain the estimates

$$N_1 = 4(y_{sk} - \sqrt{\mathcal{E}_s/2}\alpha)^2\ , \quad N_2 = 4(y_{sk} + \sqrt{\mathcal{E}_s/2}\alpha)^2 \tag{A-140}$$

Substituting these estimates into (A-137) and then substituting the results into (A-139), we obtain

$$L(y_{sk}|u_k) = \ln\left[\frac{|y_{sk} - \sqrt{\mathcal{E}_s/2}\alpha|}{|y_{sk} + \sqrt{\mathcal{E}_s/2}\alpha|}\right] \tag{A-141}$$

A *turbo block code* uses two linear block codes as its component codes. To limit the decoding complexity, high-rate binary BCH codes are generally used as the component codes, and the turbo code is called a *turbo BCH code*. The encoder of a turbo block code has the form of Figure A-17. Puncturing is generally not used as it causes a significant performance degradation. Suppose that the component block codes are binary systematic (n_1, k_1) and (n_2, k_2) codes, respectively. The interleaver has an array into which $k_1 k_2$ information bits are written as k_1 columns and k_2 rows. Encoder 2 converts each set of k_2 interleaver output bits into a codeword of n_2 bits. The multiplexer passes the n_1 bits of each of the k_2 encoder-1 codewords, but only the $n_2 - k_2$ parity bits of k_1 encoder-2 codewords so that information bits are transmitted only once. Consequently, the code rate of the turbo block code is

$$r = \frac{k_1 k_2}{k_2 n_1 + (n_2 - k_2)k_1} \tag{A-142}$$

If the two block codes are identical, then $r = k/(2n - k)$. If the minimum Hamming distances of the component codes are d_{m1} and d_{m2}, respectively, then the minimum distance of the concatenated code is

$$d_m = d_{m1} + d_{m2} - 1 \tag{A-143}$$

The decoder of a turbo block code has the form of Figure A-18, and only slight modifications of the SISO decoding algorithms are required. Long, high-rate turbo BCH codes approach the Shannon limit in performance, but their complexities are higher then those of turbo convolutional codes of comparable performance [43].

Approximate upper bounds on the bit error probability for turbo codes have been derived [36], [43]. Since these bounds are difficult to evaluate except for short codewords, simulation results are generally used to predict the performance of a turbo code.

Serial Concatenated Codes with Iterative Decoding

The encoder for a *serial concatenated code* has the form of Figure 15(a). When iterative decoding is used, usually both component codes are either binary systematic block codes or binary systematic convolutional codes. The outer encoder generates n_1 bits for every k_1 information bits. After the interleaving, each set of n_1 bits is converted by the inner encoder into n_2 bits. Thus, the overall code rate of the serial concatenated code is k_1/n_2. If the component codes are block codes, then an outer (n_1, k_1) code and an inner (n_2, n_1) code are used. A functional block diagram of an iterative decoder for a serial concatenated code is illustrated in Figure A-19. For each inner codeword, the input comprises the demodulator outputs corresponding to the n_2 bits. For each iteration, the inner decoder computes the LLRs for the n_1 systematic bits. After a deinterleaving, these LLRs provide extrinsic information about the n_1 code bits of the outer code. The outer decoder then computes the LLRs for all its code bits. After an interleaving, these LLRs provide extrinsic information about the n_1 systematic bits of the inner code. The final output of the iterative decoder comprises the k_1 information bits of the concatenated code. Simulation results indicate that a serial concatenated code with convolutional codes tends to outperform a comparable turbo convolutional code for the AWGN channel when low bit error probabilities are required [36].

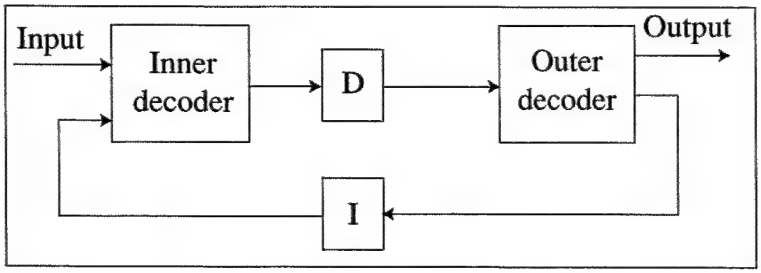


Figure A-19. Iterative decoder for serial concatenated code. D = deinterleaver; I = interleaver.

Turbo Product Codes

A *product code* is a special type of serial concatenated code that is constructed from multidimensional arrays and linear block codes. When iterative decoding is used, a product code is called a *turbo product code*. An encoder for a two-dimensional turbo product code has the form of Figure 15(a). The outer encoder produces codewords of an (n_1, k_1) code. For an inner (n_2, k_2) code, k_2 codewords are placed in a $k_2 \times n_1$ interleaver array of k_2 rows and n_1 columns. The block interleaver columns are read by the inner encoder to produce n_1 codewords of length n_2 that are transmitted. The resulting product code has

parameters $n = n_1 n_2$, $k = k_1 k_2$, and code rate

$$r = \frac{k_1 k_2}{n_1 n_2} \tag{A-144}$$

If the minimum Hamming distances of the outer and inner codes are d_{m1} and d_{m2}, respectively, then the minimum Hamming distance of the product code is

$$d_m = d_{m1} d_{m2} \tag{A-145}$$

Comparing this equation with (A-143) indicates that d_m for a turbo product code is generally larger than d_m for a turbo block code with the same component codes. The decoder for a turbo product code has the form shown in Figure A-20. The demodulator outputs are applied to both the inner decoder, and after deinterleaving, the outer decoder. The LLRs of both the information and parity bits of the corresponding code are computed by each decoder. These LLRs are then exchanged between the decoders after the appropriate deinterleaving or interleaving converts the LLRs into extrinsic information. For a given complexity, the performance of turbo product codes and turbo block codes are similar [43]. A large reduction in the complexity of a turbo product code in exchange for a relatively small performance loss is obtained by using the Chase algorithm (Section A-1) in the SISO algorithm of the component decoders [44].

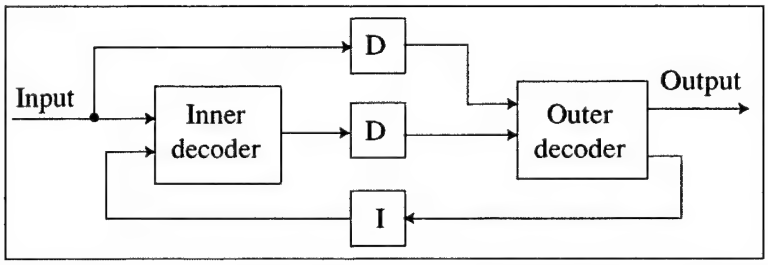


Figure A-20. Decoder of turbo product code. D = deinterleaver; I = interleaver.

Turbo Trellis-Coded Modulation

Turbo trellis-coded modulation (TTCM), which produces a nonbinary bandwidth-efficient modulation, is obtained by using identical trellis codes as the component codes in a turbo code [45]. The encoder has the form illustrated in Figure A-21. The code rate and, hence, the required bandwidth of the component trellis code is preserved by the TTCM encoder because it alternately selects constellation points or complex symbols generated by the two parallel component encoders. To ensure that all information bits, which constitute the encoder input, are transmitted only once and that the parity bits are provided alternately by the two component encoders, the symbol interleaver transfers symbols in odd positions to odd positions and symbols in even positions to even positions, where each symbol is a group of bits. After the complex symbols are produced by signal mapper 2, the symbol deinterleaver restores the original

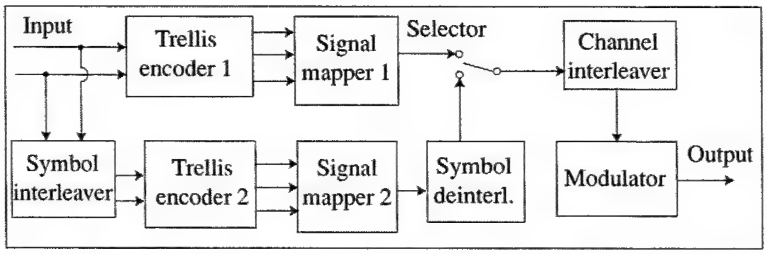


Figure A-21. Encoder for turbo trellis-coded modulation.

ordering. The selector passes the odd-numbered complex symbols from mapper 1 and the even-numbered complex symbols from mapper 2. The channel interleaver permutes the selected complex symbols prior to the modulation. The TTCM decoder uses a symbol-based SISO algorithm analogous to the SISO algorithm used by turbo-decoders. TTCM can provide a performance close to the Shannon limit for the AWGN channel, but its implementation complexity is much greater than that of conventional trellis-coded modulation [43].

Appendix B. Signal Representations

B-1 Hilbert Transform

Consider a real-valued function $g(t)$ defined in the time interval $-\infty < t < \infty$. The *Hilbert transform* of $g(t)$ is defined by

$$H[g(t)] = \hat{g}(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{g(u)}{t-u}\, du \tag{B-1}$$

Because this integrand has a singularity, we define the integral as its Cauchy principal value. Thus,

$$\int_{-\infty}^{\infty} \frac{g(u)}{t-u}\, du = \lim_{\epsilon \to 0} \left[\int_{-\infty}^{t-\epsilon} \frac{g(u)}{t-u}\, du + \int_{t+\epsilon}^{\infty} \frac{g(u)}{t-u}\, du \right] \tag{B-2}$$

provided that the limit exists. Subsequently, integrals are to be interpreted as Cauchy principal values if they contain singularities.

The definition of the Hilbert transform indicates that $\hat{g}(t)$ may be interpreted as the convolution of $g(t)$ with $1/\pi t$. Therefore, $\hat{g}(t)$ results from passing $g(t)$ through a linear filter with an impulse response equal to $1/\pi t$. The transfer function of the filter is given by the Fourier transform

$$\mathcal{F}\left\{\frac{1}{\pi t}\right\} = \int_{-\infty}^{\infty} \frac{\exp(-j2\pi f t)}{\pi t}\, dt \tag{B-3}$$

where $j = \sqrt{-1}$. This integral can be rigorously evaluated by using contour integration. Alternatively, we observe that since $1/t$ is an odd function,

$$\begin{aligned} \mathcal{F}\left\{\frac{1}{\pi t}\right\} &= -2j \int_{0}^{\infty} \frac{\sin 2\pi f t}{\pi t}\, dt \\ &= -j \operatorname{sgn}(f) \end{aligned} \tag{B-4}$$

where $\operatorname{sgn}(f)$ is the *signum function* defined by

$$\operatorname{sgn}(f) = \begin{cases} 1, & f > 0 \\ 0, & f = 0 \\ -1, & f < 0 \end{cases} \tag{B-5}$$

Let $G(f) = \mathcal{F}\{g(t)\}$, and let $\hat{G}(f) = \mathcal{F}\{\hat{g}(t)\}$. Equations (B-1) and (B-4) and the convolution theorem imply that

$$\hat{G}(f) = -j \operatorname{sgn}(f) G(f) \tag{B-6}$$

Because $H[\hat{g}(t)]$ results from passing $g(t)$ through two successive filters, each with transfer function $-j\ \mathrm{sgn}(f)$,

$$H[\hat{g}(t)] = -g(t) \tag{B-7}$$

provided that $G(0) = 0$.

Equation (B-6) indicates that taking the Hilbert transform corresponds to introducing a phase sift of -90 degrees for all positive frequencies and +90 degrees for all negative frequencies. Consequently,

$$H[\cos 2\pi f_0 t] = \sin 2\pi f_0 t \tag{B-8}$$

$$H[\sin 2\pi f_0 t] = -\cos 2\pi f_0 t \tag{B-9}$$

These relations can be formally verified by taking the Fourier transform of the left-hand side of (B-8) or (B-9), applying (B-6), and then taking the inverse Fourier transform of the result. If $G(f) = 0$ for $|f| > W$ and $f_0 > W$, the same method yields

$$H[g(t)\cos 2\pi f_0 t] = g(t)\sin 2\pi f_0 t \tag{B-10}$$

$$H[g(t)\sin 2\pi f_0 t] = -g(t)\cos 2\pi f_0 t \tag{B-11}$$

B-2 Analytic Signal and Complex Envelope

A *bandpass signal* is one with a Fourier transform that is negligible except for $f_c - W/2 \leq |f| \leq f_c + W/2$, where $0 \leq W < 2f_c$ and f_c is the center frequency. If $W << f_c$, the bandpass signal is often called a *narrowband signal.* A complex-valued signal with a Fourier transform that is nonzero only for $f > 0$ is called an *analytic signal.*

Consider a bandpass signal $g(t)$ with Fourier transform $G(f)$. The analytic signal $g_a(t)$ associated with $g(t)$ is defined to be the signal with Fourier transform

$$G_a(f) = [1 + \mathrm{sgn}(f)]G(f) \tag{B-12}$$

which is zero for $f \leq 0$ and is confined to the band $|f - f_c| \leq W/2$ when $f > 0$. The inverse Fourier transform of (B-12) and (B-6) imply that

$$g_a(t) = g(t) + j\hat{g}(t) \tag{B-13}$$

The *complex envelope* of $g(t)$ is defined by

$$g_l(t) = g_a(t)\exp[-j2\pi f_c t] \tag{B-14}$$

where f_c is the center frequency if $g(t)$ is a bandpass signal. Since the Fourier transform of $g_l(t)$ is $G_a(f + f_c)$, which occupies the band $|f| \leq W/2$, the complex envelope is a baseband signal that may be regarded as an *equivalent*

lowpass representation of $g(t)$. Equations (B-13) and (B-14) imply that $g(t)$ may be expressed in terms of its complex envelope as

$$g(t) = \mathrm{Re}[g_l(t) \exp(j2\pi f_c t)] \tag{B-15}$$

The complex envelope can be decomposed as

$$g_l(t) = g_c(t) + j g_s(t) \tag{B-16}$$

where $g_c(t)$ and $g_s(t)$ are real-valued functions. Therefore, (B-15) yields

$$g(t) = g_c(t) \cos(2\pi f_c t) - g_s(t) \sin(2\pi f_c t) \tag{B-17}$$

Since the two sinusoidal carriers are in phase quadrature, $g_c(t)$ and $g_s(t)$ are called the *in-phase* and *quadrature* components of $g(t)$, respectively. These components are lowpass signals confined to $|f| \leq W/2$.

From Parseval's relation in Fourier analysis and (B-6), we obtain

$$\int_{-\infty}^{\infty} \hat{g}^2(t)\, dt = \int_{-\infty}^{\infty} |\hat{G}(f)|^2\, df = \int_{-\infty}^{\infty} |G(f)|^2\, df = \int_{-\infty}^{\infty} g^2(t)\, dt \tag{B-18}$$

Therefore,

$$\begin{aligned} \int_{-\infty}^{\infty} |g_l(t)|^2\, dt &= \int_{-\infty}^{\infty} |g_a(t)|^2\, dt = \int_{-\infty}^{\infty} g^2(t)\, dt + \int_{-\infty}^{\infty} \hat{g}^2(t)\, dt \\ &= 2 \int_{-\infty}^{\infty} g^2(t)\, dt = 2\mathcal{E} \end{aligned} \tag{B-19}$$

where $\mathcal{E}$ denotes the energy of the bandpass signal $g(t)$.

B-3 Stationary Stochastic Processes

Consider a stochastic process $n(t)$ that is a zero-mean, wide-sense stationary process with autocorrelation

$$R_n(\tau) = E[n(t)n(t+\tau)] \tag{B-20}$$

where $E[x]$ denotes the expected value of x. The Hilbert transform of this process is the stochastic process defined by

$$\hat{n}(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{n(u)}{t-u} du \tag{B-21}$$

where it is assumed that the Cauchy principal value of the integral exists for almost every sample function of $n(t)$. This equation indicates that $\hat{n}(t)$ is a zero-mean stochastic process. The zero-mean processes $n(t)$ and $\hat{n}(t)$ are *jointly wide-sense stationary* if their correlation and cross-correlation functions are not

functions of t. A straightforward calculation using (B-21) and (B-20) gives the cross correlation

$$R_{n\hat{n}}(\tau) = E[n(t)\hat{n}(t+\tau)] = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{R_n(u)}{\tau - u}du = \hat{R}_n(\tau) \tag{B-22}$$

A similar derivation using (B-7) yields the autocorrelation

$$R_{\hat{n}}(\tau) = E[\hat{n}(t)\hat{n}(t+\tau)] = R_n(\tau) \tag{B-23}$$

Equations (B-20), (B-22), and (B-23) indicate that $n(t)$ and $\hat{n}(t)$ are jointly wide-sense stationary.

The *analytic signal* associated with $n(t)$ is the zero-mean process defined by

$$n_a(t) = n(t) + j\hat{n}(t) \tag{B-24}$$

The autocorrelation of the analytic signal is defined as

$$R_a(\tau) = E[n_a^*(t)n_a(t+\tau)] \tag{B-25}$$

where thee asterisk denotes the complex conjugate. Using (B-20) and (B-22) to (B-25), we obtain

$$R_a(\tau) = 2R_n(\tau) + 2j\hat{R}_n(\tau) \tag{B-26}$$

which establishes the wide-sense stationarity of the analytic signal.

Since (B-20) indicates that $R_n(\tau)$ is an even function, (B-22) yields

$$R_{n\hat{n}}(0) = \hat{R}_n(0) = 0 \tag{B-27}$$

which indicates that $n(t)$ and $\hat{n}(t)$ are uncorrelated. Equations (B-23), (B-26), and (B-27) yield

$$R_{\hat{n}}(0) = R_n(0) = 1/2R_a(0) \tag{B-28}$$

The *complex envelope* of $n(t)$ or the *equivalent lowpass representation* of $n(t)$ is the zero-mean stochastic process defined by

$$n_l(t) = n_a(t)\exp(-j2\pi f_c t) \tag{B-29}$$

where f_c is an arbitrary frequency usually chosen as the center or carrier frequency of $n(t)$. The complex envelope can be decomposed as

$$n_l(t) = n_c(t) + jn_s(t) \tag{B-30}$$

where $n_c(t)$ and $n_s(t)$ are real-valued, zero-mean stochastic processes.

Equations (B-29) and (B-30) imply that

$$\begin{aligned} n(t) &= \text{Re}[n_l(t)\exp(j2\pi f_c t)] \\ &= n_c(t)\cos(2\pi f_c t) - n_s(t)\sin(2\pi f_c t) \end{aligned} \tag{B-31}$$

Substituting (B-24) and (B-30) into (B-29) we find that

$$n_c(t) = n(t)\cos(2\pi f_c t) + \hat{n}(t)\sin(2\pi f_c t) \tag{B-32}$$

$$n_s(t) = \hat{n}(t)\cos(2\pi f_c t) - n(t)\sin(2\pi f_c t) \tag{B-33}$$

The *autocorrelations* of $n_c(t)$ and $n_s(t)$ are defined by

$$R_c(\tau) = E[n_c(t)n_c(t+\tau)] \tag{B-34}$$

and

$$R_s(\tau) = E[n_s(t)n_s(t+\tau)] \tag{B-35}$$

Using (B-32) and (B-33) and then (B-20), (B-23), and (B-24) and trigonometric identities, we obtain

$$R_c(\tau) = R_s(\tau) = R_n(\tau)\cos(2\pi f_c \tau) + \hat{R}_n(\tau)\sin(2\pi f_c \tau) \tag{B-36}$$

which shows explicitly that if $n(t)$ is wide-sense stationary, then $n_c(t)$ and $n_s(t)$ are wide-sense stationary with the same autocorrelation function. The variances of $n(t)$, $n_c(t)$, and $n_s(t)$ are all equal because

$$R_c(0) = R_s(0) = R_n(0) \tag{B-37}$$

A derivation similar to that of (B-36) gives the cross correlation

$$R_{cs}(\tau) = E[n_c(t)n_s(t+\tau)] = \hat{R}_n(\tau)\cos(2\pi f_c \tau) - R_n(\tau)\sin(2\pi f_c \tau) \tag{B-38}$$

Equations (B-36) and (B-38) indicate that $n_c(t)$ and $n_s(t)$ are jointly wide-sense stationary. Equations (B-28) and (B-38) give

$$R_{cs}(0) = 0 \tag{B-39}$$

which implies that $n_c(t)$ and $n_s(t)$ are uncorrelated.

Equation (B-21) indicates that $\hat{n}(t)$ is generated by a linear operation on $n(t)$. Therefore, if $n(t)$ is a zero-mean Gaussian process, $\hat{n}(t)$ and $n(t)$ are zero-mean jointly Gaussian processes. Equations (B-32) and (B-33) then imply that $n_c(t)$ and $n_s(t)$ are zero-mean jointly Gaussian processes. Since they are uncorrelated, $n_c(t)$ and $n_s(t)$ are statistically independent, zero-mean Gaussian processes.

The *power spectral density* of a signal is the Fourier transform of its autocorrelation. Let $S(f)$, $S_c(f)$, and $S_s(f)$ denote the power spectral densities of $n(t)$, $n_c(t)$, and $n_s(t)$, respectively. We assume that $S_n(f)$ occupies the band

$f_c - W/2 \leq |f| \leq f_c + W/2$ and that $f_c > W/2 \geq 0$. Taking the Fourier transform of (B-36), using (B-6), and simplifying, we obtain

$$S_c(f) = S_s(f) = \begin{cases} S_n(f - f_c) + S_n(f + f_c), & |f| \leq W/2 \\ 0, & |f| > W/2 \end{cases} \tag{B-40}$$

Thus, if $n(t)$ is a passband process with one-sided bandwidth W, then $n_c(t)$ and $n_s(t)$ are baseband processes with one-sided bandwidths $W/2$. This property and the statistical independence of $n_c(t)$ and $n_s(t)$ when $n(t)$ is Gaussian make (B-31) a very useful representation of $n(t)$.

Similarly, the cross-spectral density of $n_c(t)$ and $n_s(t)$ can be derived by taking the Fourier transform of (B-38) and using (B-6). After simplification, the result is

$$S_{cs}(f) = \begin{cases} j[S_n(f - f_c) - S_n(f + f_c)], & |f| \leq W/2 \\ 0, & |f| > W/2 \end{cases} \tag{B-41}$$

If $S_n(f)$ is locally symmetric about f_c, then

$$S_n(f_c + f) = S_n(f_c - f), \quad |f| \leq W/2 \tag{B-42}$$

Since a power spectral density is a real-valued, even function, $S_n(f_c - f) = S_n(f - f_c)$. Equation (B-42) then yields $S_n(f + f_c) = S_n(f - f_c)$ for $|f| \leq W/2$. Therefore, (B-41) gives $S_{cs}(f) = 0$, which implies that

$$R_{cs}(\tau) = 0 \tag{B-43}$$

for all τ. Thus, $n_c(t)$ and $n_s(t + \tau)$ are uncorrelated for all τ, and if $n(t)$ is a zero-mean Gaussian process, then $n_c(t)$ and $n_s(t + \tau)$ are statistically independent for all τ.

The autocorrelation of the complex envelope is defined by

$$R_l(\tau) = \frac{1}{2} E[n_l^*(t) n_l(t + \tau)] \tag{B-44}$$

where the 1/2 is inserted so that

$$R_l(0) = R_n(0) \tag{B-45}$$

which follows from (B-28) and (B-29). Substituting (B-30) into (B-44) and using (B-36) and (B-38), we obtain

$$R_l(\tau) = R_c(\tau) + jR_{cs}(\tau) \tag{B-46}$$

The power spectral density of $n_l(t)$, which we denote by $S_l(f)$, can be derived from (B-46), (B-41), and (B-40). If $S_n(f)$ occupies the band $f_c - W/2 \leq |f| \leq f_c + W/2$ and $f_c > W/2 \geq 0$, then

$$S_l(f) = \begin{cases} 2S_n(f + f_c), & |f| \leq W/2 \\ 0, & |f| > W/2 \end{cases} \tag{B-47}$$

Equations (B-36) and (B-38) yield

$$R_n(\tau) = R_c(\tau)\cos(2\pi f_c\tau) - R_{cs}(\tau)\sin(2\pi f_c\tau) \tag{B-48}$$

Equations (B-48) and (B-46) imply that

$$R_n(\tau) = \text{Re}\left[R_l(\tau)\exp\left(j2\pi f_c\tau\right)\right] \tag{B-49}$$

We expand the right-hand side of this equation by using the fact that $\text{Re}[z] = (z + z^*)/2$. Taking the Fourier transform and observing that $S_l(f)$ is a real-valued function, we obtain

$$S_n(f) = \frac{1}{2}S_l(f - f_c) + \frac{1}{2}S_l(-f - f_c) \tag{B-50}$$

If $S_n(f)$ is locally symmetric about f_c, then (B-47) and (B-42) imply that $S_l(-f) = S_l(f)$, and (B-50) becomes

$$S_n(f) = \frac{1}{2}S_l(f - f_c) + \frac{1}{2}S_l(f + f_c) \tag{B-51}$$

Many useful communication signals are modeled as having the form

$$s(t) = Ad_1(t)\cos(2\pi f_c t + \theta) + Ad_2(t)\sin(2\pi f_c t + \theta) \tag{B-52}$$

where θ is an independent random variable that is uniformly distributed over $0 \leq \theta < 2\pi$. The modulations have the form

$$d_i(t) = \sum_{k=-\infty}^{\infty} a_{ik}\psi(t - kT - T_0 - t_i), \quad i = 1, 2 \tag{B-53}$$

where $\{a_{ik}\}$ is a sequence of independent, identically distributed random variables, $a_{ik} = +1$ with probability 1/2 and $a_{ik} = -1$ with probability 1/2, $\psi(t)$ is a pulse waveform, T is the pulse duration, t_i is the relative pulse offset, and T_0 is an independent random variable that is uniformly distributed over the interval $(0, T)$ and reflects the arbitrariness of the origin of the coordinate system. Since a_{ik} is independent of a_{in} when $n \neq k$, it follows that $E[a_{ik}a_{in}] = 0, n \neq k$. It follows that the autocorrelation of $d_i(t)$ is

$$\begin{aligned} R_{di}(\tau) &= E[d_i(t)d_i(t + \tau)] \\ &= \sum_{k=-\infty}^{\infty} E[\psi(t - kT - T_0 - t_i)\psi(t - kT - T_0 - t_i + \tau)] \end{aligned} \tag{B-54}$$

Expressing the expected value as an integral over the range of T_0 and changing variables, we obtain

$$\begin{aligned} R_{di}(\tau) &= \sum_{k=-\infty}^{\infty} \frac{1}{T}\int_{t-kT-T-t_i}^{t-kT-t_i} \psi(x)\psi(x + \tau)dx \\ &= \frac{1}{T}\int_{-\infty}^{\infty} \psi(x)\psi(x + \tau)dx, \quad i = 1, 2 \end{aligned} \tag{B-55}$$

This equation indicates that $d_1(t)$ and $d_2(t)$ are wide-sense stationary processes with the same autocorrelation.

If the sequences $\{a_{1k}\}$ and $\{a_{2k}\}$ are statistically independent, then the autocorrelation of $s(t)$ is

$$R_s(\tau) = \frac{A^2}{2} R_{d1}(\tau) \cos(2\pi f_c \tau) + \frac{A^2}{2} R_{d2}(\tau) \sin(2\pi f_c \tau) \tag{B-56}$$

where $R_{d1}(\tau)$ and $R_{d2}(\tau)$ are the autocorrelations of $d_1(t)$ and $d_2(t)$, respectively. This equation indicates that $s(t)$ is wide-sense stationary. If the sample functions of $d_1(t)$ and $d_2(t)$ have Fourier transforms that vanish for $|f| \geq f_c$, then (B-10), (B-11), (B-24), and (B-29) indicate that the complex envelope of $s(t)$ is

$$s_l(t) = Ad_1(t) - jAd_2(t) \tag{B-57}$$

Equation (B-44) and the independence of $d_1(t)$ and $d_2(t)$ imply that the autocorrelation of $s_l(t)$ is

$$R_l(\tau) = \frac{A^2}{2} R_{d1}(\tau) + \frac{A^2}{2} R_{d2}(\tau) \tag{B-58}$$

The power spectral density of $s_l(t)$ is the Fourier transform of $R_l(\tau)$. From (B-58) and (B-55), we obtain the density

$$S_l(f) = A^2 \frac{|G(f)|^2}{T} \tag{B-59}$$

where $G(f)$ is the Fourier transform of $\psi(t)$.

In a quadriphase-shift-keying (QPSK) signal, $d_1(t)$ and $d_2(t)$ are usually modeled as independent random binary sequences with pulse duration $T = 2T_b$, where T_b is a bit duration. The component amplitude is $A = \sqrt{\mathcal{E}_b/T_b}$, where $\mathcal{E}_b$ is the energy per bit. If $\psi(t)$ is rectangular with unit amplitude over $[0, 2T_b]$, then (B-59) yields the power spectral density for QPSK:

$$S_l(f) = 2\mathcal{E}_b \text{sinc}^2 2T_b f \tag{B-60}$$

which is the same as the density for PSK. For a binary minimum-shift-keying (MSK) signal with the same component amplitude,

$$\psi(t) = \sqrt{2} \sin\left(\frac{\pi t}{2T_b}\right), \quad 0 \leq t < 2T_b \tag{B-61}$$

Therefore, the power spectral density for MSK is

$$S_l(f) = \frac{16\mathcal{E}_b}{\pi^2} \left[\frac{\cos(2\pi T_b f)}{16T_b^2 f^2 - 1}\right]^2 \tag{B-62}$$

B-4 Sampling Theorems

Consider the Fourier transform $G(f)$ of an absolutely integrable function $g(t)$. The periodic extension of $G(f)$ is defined as

$$\bar{G}(f) = \sum_{i=-\infty}^{\infty} G(f + iW) \tag{B-63}$$

where W is the period of $\bar{G}(f)$ and it is assumed that the series converges uniformly. Suppose that $\bar{G}(f)$ has a piecewise continuous derivative so that it can be represented as a uniformly convergent complex Fourier series:

$$\bar{G}(f) = \sum_{k=-\infty}^{\infty} c_k \exp\left(-j2\pi k \frac{f}{W}\right) \tag{B-64}$$

where the Fourier coefficient c_k is given by

$$c_k = \frac{1}{W} \int_{-W/2}^{W/2} \bar{G}(f) \exp\left(j2\pi k \frac{f}{W}\right) \, df \tag{B-65}$$

Substituting (B-63) into (B-65) and interchanging the order of the summation and the integration, which is justified because of the uniform convergence, we obtain

$$c_k = \frac{1}{W} \sum_{i=-\infty}^{\infty} \int_{-W/2}^{W/2} G(f + iW) \exp\left(j2\pi k \frac{f}{W}\right) \, df \tag{B-66}$$

We change variables and observe the $\exp(j2\pi ki) = 1$ to obtain

$$\begin{aligned} c_k &= \frac{1}{W} \sum_{i=-\infty}^{\infty} \int_{-W/2+iW}^{W/2+iW} G(f) \exp\left(j2\pi k \frac{f}{W} - j2\pi ki\right) \, df \\ &= \frac{1}{W} \int_{-\infty}^{\infty} G(f) \exp\left(j2\pi k \frac{f}{W}\right) \, df \end{aligned} \tag{B-67}$$

Since $g(t)$ is absolutely integrable, the last integral is the inverse Fourier transform of $G(f)$ evaluated at $t = k/W$, and

$$c_k = \frac{1}{W} g\left(\frac{k}{W}\right) \tag{B-68}$$

Substituting (B-68) into (B-64) yields the *Poisson sum formula*:

$$\bar{G}(f) = \frac{1}{W} \sum_{k=-\infty}^{\infty} g\left(\frac{k}{W}\right) \exp\left(-\frac{j2\pi kf}{W}\right) \tag{B-69}$$

where the series converges uniformly.

Suppose that the Fourier transform vanishes outside a frequency band:

$$G(f) = 0\,, \quad |f| > W/2 \tag{B-70}$$

It follows that

$$g(t) = \int_{-W/2}^{W/2} G(f) \exp(j2\pi ft) df \tag{B-71}$$

Since $G(f) = \bar{G}(f)$ for $|f| < W/2$, (B-71) and (B-69) and the interchange of a summation and integration yield

$$g(t) = \sum_{k=-\infty}^{\infty} g\left(\frac{k}{W}\right) \frac{1}{W} \int_{-W/2}^{W/2} \exp\left[j2\pi f\left(t - \frac{k}{W}\right)\right] df \tag{B-72}$$

Evaluating this integral and defining

$$\text{sinc } x = \frac{\sin \pi x}{\pi x} \tag{B-73}$$

we obtain the *sampling theorem* for deterministic signals:

$$g(t) = \sum_{k=-\infty}^{\infty} g\left(\frac{k}{W}\right) \text{sinc}(Wt - k) \tag{B-74}$$

Consider a wide-sense stationary stochastic process $n(t)$ with autocorrelation $R_n(\tau)$ and power spectral density $S_n(f)$, which is the Fourier transform of $R_n(\tau)$. If

$$S_n(f) = 0\,, \quad |f| > W/2 \tag{B-75}$$

then it follows from the sampling theorem that

$$R_n(\tau) = \sum_{k=-\infty}^{\infty} R_n\left(\frac{k}{W}\right) \text{sinc}(W\tau - k) \tag{B-76}$$

For an arbitrary constant α, the Fourier transform of $R(\tau - \alpha)$ is $S_n(f)$ $\exp(-j2\pi f\alpha)$, which is zero for $|f| > W/2$. Therefore, (B-76) can be applied to $R'_n(\tau) = R_n(\tau - \alpha)$, which gives

$$R_n(\tau - \alpha) = \sum_{k=-\infty}^{\infty} R_n\left(\frac{k}{W} - \alpha\right) \text{sinc}(W\tau - k) \tag{B-77}$$

We define the stochastic process

$$n_\nu(t) = \sum_{k=-\nu}^{\nu} n\left(\frac{k}{W}\right) \text{sinc}(Wt - k) \tag{B-78}$$

An expansion indicates that the mean square difference between $n(t)$ and $n_\nu(t)$ is

$$E\{[n(t) - n_\nu(t)]^2\} = R_n(0) - 2\sum_{k=-\nu}^{\nu} R_n\left(t - \frac{k}{W}\right)\text{sinc}(Wt - k) + \sum_{i=-\nu}^{\nu}\text{sinc}(Wt - i)\sum_{k=-\nu}^{\nu} R_n\left(\frac{i-k}{W}\right)\text{sinc}(Wt - k) \quad \text{(B-79)}$$

Since $R_n(\tau) = R_n(-\tau)$, the repeated use of (B-77) yields

$$\lim_{\nu\to\infty} E\{[n(t) - n_\nu(t)]^2\} = 0 \quad \text{(B-80)}$$

which states that the mean square difference between $n(t)$ and $n_\nu(t)$ approaches zero. Thus, the sampling theorem for stationary stochastic process is

$$n(t) = \sum_{k=-\infty}^{\infty} n\left(\frac{k}{W}\right)\text{sinc}(Wt - k) \quad \text{(B-81)}$$

where the equality holds in the sense of (B-80).

References

1. P. V. Kumar, "Frequency-Hopping Code Sequence Designs Having Large Linear Span," *IEEE Trans. Inform. Theory*, **34**, pp. 146–151, January 1988.

2. L. Cong and S. Songgeng, "Chaotic Frequency Hopping Sequences," *IEEE Trans. Commun*, **46**, pp. 1433–1437, November 1998.

3. J. R. Smith, *Modern Communication Circuits, 2nd ed.* Boston: McGraw-Hill, 1998.

4. U. L. Rohde, *Microwave and Wireless Synthesizers, Theory and Design.* New York: Wiley, 1997.

5. W. F. Egan, *Frequency Synthesis by Phase Lock, 2nd ed.* New York: Wiley, 2000.

6. J. R. Alexovich and R. M. Gagliardi, "Effect of PLL Frequency Synthesizer in FSK Frequency-Hopped Communications," *IEEE Trans. Commun.*, **37**, pp. 268–276, March 1989.

7. D. J. Torrieri "Fundamental Limitations on Repeater Jamming of Frequency-Hopping Communications," *IEEE J. Select. Areas Commun.*, **7**, pp. 569–578, May 1989.

8. R. L. Peterson, R. E. Ziemer, and D. E. Borth, *Introduction to Spread Spectrum Communications*. Upper Saddle River, NJ: Prentice Hall, 1995.

9. M. K. Simon, B. K. Levitt, J. K. Omura, and R. A. Scholtz, *Spread Spectrum Communications Handbook*. Boston: McGraw-Hill, 1994.

10. D. J. Torrieri, "Frequency Hopping with Multiple Frequency-Shift Keying and Hard Decisions," *IEEE Trans. Commun.*, **32**, pp. 574–583, May 1984.

11. J. S. Lee, R. H. French, and L. E. Miller, "Probability of Error Analyses of a BFSK Frequency-Hopping System with Diversity under Partial-Band Jamming Interference—Part I: Performance of Square-Law Linear Combining Soft Decision Receiver," *IEEE Trans. Commun.*, **32**, pp. 645–653, June 1984.

12. J. S. Lee, L. E. Miller, and Y. K. Kim, "Probability of Error Analyses of a BFSK Frequency-Hopping System with Diversity under Partial-Band Jamming Interference—Part II: Performance of Square-Law Nonlinear Combining Soft Decision Receivers," *IEEE Trans. Commun.*, **32**, pp. 1243–1250, December 1984.

13. L. E. Miller, J. S. Lee, and A. P. Kadrichu, "Probability of Error Analyses of a BFSK Frequency-Hopping System with Diversity under Partial-Band Jamming Interference—Part III: Performance of Square-Law Self-Normalizing Soft Decision Receiver," *IEEE Trans. Commun.*, **34**, pp. 669–675, July 1986.

14. Y. M. Lam and P. H. Wittke, "Frequency-Hopped Spread-Spectrum Transmission with Band-Efficient Modulations and Simplified Noncoherent Sequence Estimation," *IEEE Trans. Commun.*, **38**, pp. 2184–2196, December 1990.

15. M. K. Simon, S. M. Hinedi, and W. C. Lindsey, *Digital Communication Techniques*. Englewood Cliffs, NJ: Prentice Hall, 1995.

16. R. F. Pawula, "On the Theory of Error Rates for Narrow-Band Digital FM," *IEEE Trans. Commun.*, **29**, pp. 1634–1643, November 1981.

17. M. B. Pursley, "The Derivation and Use of Side Information in Frequency-Hop Spread Spectrum Communications," *IEICE Trans. Commun.*, **E76-B**, pp. 814–824, August 1993.

18. L.-L. Yang and L. Hanzo, "Low Complexity Erasure Insertion in RS-Coded SFH Spread-Sprectum Communications With Partial-Band Interference and Nakagami-m Fading," *IEEE Trans. Commun.*, **50**, pp. 914–925, June 2002.

19. P. H. Wittke, Y. M. Lam, and M. J. Schefter, "The Performance of Trellis-Coded Nonorthogonal Noncoherent FSK in Noise and Jamming," *IEEE Trans. Commun.*, **43**, pp. 635–645, February/March/April 1995.

20. H. El Gamal and E. Geraniotis, "Iterative Chanel Estimation and Decoding for Convolutionally Coded Anti-Jam FH Signals," *IEEE Trans. Commun.*, **50**, pp. 321–331, February 2002.

21. J. H. Kang and W. E. Stark, "Iterative Estimation and Decoding for FH-SS with Slow Rayleigh Fading," *IEEE Trans. Commun.*, **48**, pp. 2014–2023, December 2000.

22. D J. Torrieri, "Mobile Frequency-Hopping CDMA Systems," *IEEE Trans. Commun.*, **48**, pp. 1318–1327, August 2000.

23. F. Adachi and J. D. Parsons, "Unified Analysis of Postdetection Diversity for Binary Digital FM Radio," *IEEE Trans. Veh. Technol.*, **37**, pp. 189–198, November, 1988.

24. D. Torrieri, *Principles of Secure Communication Systems, 2nd ed.* Boston: Artech House, 1992.

25. A. A. Shaar and P. A. Davies, "Survey of One-Coincidence Sequences for Frequency Hopped Spread Spectrum Systems," *Proc. IEE, Pt. F, Commun., Radar Signal Processing*, pp. 719–724, December 1984.

26. S. Chennakeshu et al., "Capacity Analysis of a TDMA-Based Slow-Frequency-Hopped Cellular System," *IEEE Trans. Veh. Technol.*, **45**, pp. 531–542, August 1996.

27. L. E. Miller, J. S. Lee, R. H. French, and D. J. Torrieri, "Analysis of an Anti-jam FH Acquisition Scheme," *IEEE Trans. Commun.*, **40**, pp. 160–170, January 1992.

28. C. A. Putman, S. S. Rappaport, and D. L. Schilling, "Comparison of Strategies for Serial Acquisition of Frequency-Hopped Spread-Spectrum Signals," *IEE Proc.*, F, **133**, pp. 129–137, April 1986.

29. C. A. Putman, S. S. Rappaport, and D. L. Schilling, "Tracking of Frequency-Hopped Spread-Spectrum Signals in Adverse Environments," *IEEE Trans. Commun.*, **31**, pp. 955–963, August 1983.

30. J. H. Gass, D. L. Noneaker, and M. B. Pursley, "A Comparison of Slow-Frequency-Hop and Direct-Sequence Spread-Spectrum Packet Communications Over Doubly Selective Fading Channels," *IEEE Trans. Commun.*, **50**, pp. 1236–1239, August 2002.

31. M. D. Srinath, P. K. Rajasekaran, and R. Viswanathan, *Introduction to Statistical Signal Processing with Applications*. Englewood Cliffs, NJ: Prentice Hall, 1996.

32. B. K. Levitt et al., "Optimum Detection of Slow Frequency-Hopped Signals," *IEEE Trans. Commun.*, **42**, pp. 1990–2000, February/ March/April 1994.

33. N. C. Beaulieu, W. L. Hopkins, and P. J. McLane, "Interception of Frequency-Hopped Spread-Spectrum Signals," *IEEE. J. Select. Areas Commun.*, **8**, pp. 853–870, June 1990.

34. D. A. Hill and E. B. Felstead, "Laboratory Performance of Spread Spectum Detectors," *Proc. IEE-Commun.*, **142**, pp. 243–249, August 1995.

35. L. E. Miller, J. S. Lee, and D. J. Torrieri, "Frequency-Hopping Signal Detection Using Partial Band Coverage," *IEEE Trans. Aerosp. Electron. Syst.*, **29**, pp. 540–553, April 1993.

36. S. Benedetto and E. Biglieri, *Principles of Digital Transmission*. New York: Kluwer Academic, 1999.

37. S. B. Wicker, *Error Control Systems for Digital Communication and Storage*. Upper Saddle River, NJ: Prentice-Hall, 1995.

38. S. G. Wilson, *Digital Modulation and Coding*. Upper Saddle River, NJ: Prentice-Hall, 1996.

39. J. G. Proakis, *Digital Communications, 4th ed.* New York: McGraw-Hill, 2001.

40. D. Torrieri, "Information-Bit, Information-Symbol, and Decoded Symbol Error Rates for Linear Block Codes," *IEEE Trans. Commun.*, **36**, pp. 613–617, May 1988.

41. J.-J. Chang, D.-J. Hwang, and M.-C. Lin, "Some Extended Results on the Search for Good Convolutional Codes," *IEEE Trans. Inform. Theory*, **43**, pp. 1682–1697, September 1997.

42. C. Berrou and A. Glavieux, "Near Optimum Error-Correcting Coding and Decoding: Turbo Codes," *IEEE Trans. Commun.*, **44**, pp. 1261–1271, October 1996.

43. L. Hanzo, T. H. Liew, and B. L. Yeap, *Turbo Coding, Turbo Equalisation and Space-Time Coding*. Chichester, England: Wiley, 2002.

44. R. Pyndiah, "Near-Optimum Decoding of Product Codes: Block Turbo Codes," *IEEE Trans. Commun.*, **46**, pp. 1003–1010, August 1998.

45. P. Robertson and T. Worz, "Bandwidth Efficient Turbo Trellis-Coded Modulation Using Punctured Component Codes," *IEEE J. Selected Areas Commun.*, **16**, pp. 206–218, February 1998.